Communities to Clean Data Source Comparison

# Abstract

The primary focus of the Communities to Clean project is presenting community leaders with easily accessible data that can enable them to site renewable energy projects without an extensive monetary commitment. In order for this to be successful, users need to have access to a variety of data sources that can present a comprehensive view of a certain site. As a result, this report aims to offer an in-depth comparison of the possible data sources, in order to justify their inclusion in the project and clearly present their strengths and limitations.

The data sources that will be compared in this report are NASA POWER, the Department of Energy’s Wind Toolkit, Open Weather, the National Weather Service’s API, and meteorologic tower data from the Alaska Energy Authority. These sources were each selected because they offered an advantage not present in the others. Although they each have downfalls, the hope of this project is that through the presentation of multiple public data sources, with clearly analyzed assumptions, renewable energy siting can be done significantly quicker and cheaper.

# Data

## NASA POWER

NASA’s Prediction Of Worldwide Energy Resources (POWER) aims to provide public access to predictive meteorologic data[[1]](#footnote-1). POWER leverages NASA’s extensive data collection and processing abilities to extrapolate meteorologic information for any point on earth. This makes it an incredibly versatile data source, that can help intimate information for more remote locations that other data sources may miss.

It offers a spatial resolution of .5° x .625°, which is incredibly low. Depending on the location, meteorologic tower data can often be found within this range, which has the benefit of being empirically collected data. However, the advantage of the NASA data is that its vertical interpolation, to account for hub height, is done within a consistent model, where meteorologic tower data would require its own interpolation.

Its temporal resolution is an hourly average, which is standard for these data sources and has data from “1981 to within several months of real time.”[[2]](#footnote-2) This means that it has the greatest data availability of any of these sources, largely thanks to the strength of NASA’s predictions. The other major advantage of this source is that it comes with a built-in corrected wind speed model.[[3]](#footnote-3) Users can input the terrain type of the site that they are considering, and NASA will use this information to more accurately predict how the wind speed will change with hub height. This is significantly more accurate than the other vertical interpolation models that will be used, so it is a huge advantage of this data source.

Overall NASA POWER is the most comprehensive of these sources, offering a large variety of data, with the major caveat that little to none of it is empirically collected. The accuracy is entirely dependent on NASA’s models and the undisclosed availability of information that they have.

## Wind Toolkit

The Department of Energy’s Wind Toolkit operates in a similar manner, instead using data from wind turbines to source more empirical power information.[[4]](#footnote-4) The extrapolations for this source are slightly more limited, likely to account for the greater accuracy it can provide. It only has data for the continental United States and is incredibly limited in Alaska (citation needed).

However, these limitations mean that it can provide the highest spatial and temporal resolution. It has a spatial resolution of 2km x 2km (or about 0.018° x 0.018°), and a temporal resolution of down to a 5-minute average, which is something that none of the other data sources can offer. This is somewhat limited by the data only being available from 2007-2013, but still provides incredibly specific information if it’s required.

Its vertical interpolation is based on a different model than NASA’s Corrected Wind Speed and doesn’t account for terrain type. This means that it could be slightly less accurate, but for higher hub heights, it should be reasonably irrelevant. It provides information for the discrete hub heights of 10, 40, 60, 80, 100, 120, 140, 160 and 200m, but if a more specific value was required, it could be done with a simple calculation.

For data limited to the US, Wind Toolkit is one of the most comprehensive sources, with some small technical limitations. Due to the large nature of the data, instead of responding to an API call with a JSON or csv, a download link is provided with an email being sent to the requester when the data is finished being analyzed. While this isn’t the largest obstacle in aggregating this data, it makes it significantly harder to present in a front-end web application and will make it significantly more inaccessible to community leaders. On top of this, each API call should provide all required data in one call, which means that rate limiting isn’t a major issue. The caveat here is that the request size is limited instead by the defined request weight.

Equation 1 Request Weight, Wind Toolkit allows for a max request weight of 175,000,000. Site count is based on the amount of sites containing data for a certain point. Attribute count is the number of data points requested. Year count is how many years of data were requested. Data intervals per year is how many points of data are being requested every year.

If used properly, Wind Toolkit is a strong dataset for siting renewable energy projects in the US. It has the same limitations as other model-based data sources, but the advantage of being significantly more accurate.

## *Open Weather*

Open Weather is an Open-Source alternative to Weather Underground, both platforms focus on the collection of data from Personal Weather Stations (PWS) and meteorological extrapolation methods to provide accurate localized weather data.

There are notable benefits to this type of data. Primarily, the temporal resolution goes down to a minute, so it is the only service that can provide near real-time data. It also has 40+ years of historical data thanks to its extrapolation models. On top of this the spatial resolution is exact.

However, these benefits come with some significant downfalls. First, this falls into the same category as NASA POWER, since it is heavily relying on extrapolated data. The primary difference is that POWER has the advantage of being designed specifically for these metrics, with corrected wind speed parameters, that Open Weather does not have. This means that vertical interpolation would have to be done after Open Weather’s model has done its work, further distorting the raw data. Open Weather’s main benefit, and reason for inclusion, is the sheer number of data sources. Open Weather has 80,000 personal weather stations that it draws information from, and this can make it an incredibly versatile source.[[5]](#footnote-5) Although there is no geographic information provided about these stations (which side of a house it’s on) it can still provide more reliable information where other sources have less availability.

Another important aspect to note is that the API requires a call for each time step. On top of this, a user key has only 1,000 free calls per day. If Open Weather will be included in the platform, it will be essential to discuss how this implementation can work best for all stakeholders.

Open Weather is not a perfect source, but it is an essential one, providing information where it may not be available from others, at the cost of some accuracy.

## National Weather Service

The National Weather Service (NWS) API provides access to meteorologic data from all NWS stations. When a point is provided, raw data from any of the stations within the area is returned.

The NWS does not use any extrapolation models, so the spatial resolution is based entirely on the proximity of a station. The use of empirical station data also means that there is some variance in data availability. The temporal resolution seems to vary from station to station, with all of them at least being less than an hour. One of the major downfalls of this data source is that the only data publicly available is very close to the current date. This makes it incredibly hard to compare to the other data sources, but possibly makes it more valuable for getting a better understanding of current data than sources like Alaska Energy Authority and Wind Toolkit whose data is near 10 years old at this point.

The fact that there is only about a month of data available at any time is a severe downfall and the lack of any vertical interpolation model means that it must be done manually, but having up-to-date information is still incredibly important. This data could also possibly be archived to create an accurate and modern long-term data source that isn’t available elsewhere. Overall, despite its shortcomings this is a data source worth including.

## Alaska Energy Authority

Perhaps the most limited in scope, this data source still offers incredibly potent information by focusing on an incredibly valuable area in renewable energy that has a serious lack of data availability in other sources. Wind Toolkit is a very helpful tool for the mainland US but has lots of missing information in Alaska. The Alaska Energy Authority has no established API, but one was developed to enable easier data access for this project.

This information is again a station-based data source which means that the spatial resolution is exact, but reliant on where stations are located. One of the most significant downfalls, however, is its temporal resolution. At best, (when analyzing raw data) data is recorded hourly. Most of the data presented is more summary based and even less precise than that. On top of this, all of the data available ranges from the 1970s to the early 2000s, there is no modern data available from this source.

This makes it slightly less useful for actual data use but provides a potent comparison for the rest of the data sources. The Alaska Energy Authority acts as a keystone for the data comparison in this report as it provides the ability to compare the model-based data sources to an empirical data source in an area that historically has less data availability.

# Methods

1. “NASA POWER | Prediction Of Worldwide Energy Resources.” *NASA*, https://power.larc.nasa.gov/. Accessed 21 June 2022.

   [↑](#footnote-ref-1)
2. NASA POWER | Docs | Methodology - NASA POWER | Docs. https://power.larc.nasa.gov/docs/methodology/. Accessed 21 June 2022. [↑](#footnote-ref-2)
3. Wind Speed - NASA POWER | Docs. https://power.larc.nasa.gov/docs/methodology/meteorology/wind/. Accessed 21 June 2022. [↑](#footnote-ref-3)
4. Draxl, Caroline, et al. “The Wind Integration National Dataset (WIND) Toolkit.” Applied Energy, vol. 151, Aug. 2015, pp. 355–66. ScienceDirect, https://doi.org/10.1016/j.apenergy.2015.03.121. [↑](#footnote-ref-4)
5. “Сurrent Weather and Forecast - OpenWeatherMap.” OpenWeather, https://openweathermap.org/. Accessed 5 July 2022. [↑](#footnote-ref-5)