



Weather Datasets for Power System Modeling in India

QUESTIONS AND ANSWERS

This document was prepared to answer common questions on the weather and renewable datasets that a prospective user may have. It is not intended to be comprehensive but hopefully provides some additional context and details beyond what has been shared in the data description documents. It is organized along the categories of weather datasets, renewable datasets and power system modeling, and mechanics of data access. This document was prepared by GridLab, RMI, National Renewable Energy Laboratory (NREL), Vaisala and Electric Power Research Institute (EPRI).

Broad modeling themes

Why are weather and renewable data sets important for power system modeling (PSM)?

As we build more renewables onto the grid, it's important to understand how their production will vary with changing weather conditions, from year to year and within a year. For example, are there low renewable energy (RE) periods that we may anticipate? Because we have not built RE on all possible locations yet we don't have representative observation data for all potential generation areas. Also, where data is collected at existing projects, plant owners are typically reluctant to share it, and where it is made available the length of the record is rarely long enough for PSM use. Thus, we have to use "modeled" or simulated weather data, which is converted into simulated renewable energy production data.

How can I use these datasets, other than for capacity expansion modeling or production cost modeling?

The datasets are useful to understand how weather and possible renewable energy production varies across a particular year and from year to year across India without having to run a power system model.



How can we account for the fact that simulated data are never the same as observational data? How is the study going to address these inaccuracies?

This is a common issue with simulated data. A common approach to rectify this is to adjust the renewable profiles according to a known bias. We are in the process of conducting some validation exercises and will adjust accordingly when conducting the capacity expansion and production cost modeling steps.

Will the processed datasets be redone after the validation effort is completed?

No, we do not intend to redevelop the processed datasets that have been shared.

How can past weather data be used in capacity expansion modeling? Wouldn't there be variation in solar PV and wind capacity factors due to climate change?

The variability of today's climate can be broken into two components: natural variability and the impact of climate change. Natural variability is something we can analyze in detail today. The impact of climate change is something we can only extrapolate using existing trends and modeling efforts to make a best guess. We have very good data associated with past atmospheric conditions, particularly since the beginning of the satellite age. This allows us to look at the combined natural variability and climate change signal with a reasonable amount of accuracy using past observations, and to extrapolate that to higher spatial and temporal resolution using models which can be validated with actual data. None of this is possible with future data.

Weather datasets

Can we use AI based datasets for weather, instead of these datasets?

AI based datasets are very early in maturity. Fundamentally, good AI datasets require good calibration with physics based datasets, which can come from observation or simulations. As noted, since we have a dearth of observational RE data, good AI based weather datasets will need good simulation datasets. This initiative is a step in that direction but not sufficient.

How will the datasets, especially solar, account for air pollution and clouds?

The solar data approach, which is based on satellite measurements, can directly "see" clouds. The algorithm also uses information from global modelling centers (NASA MERRA2) regarding aerosols in the atmosphere. Therefore the approach to create solar data we used for this work explicitly accounts for air pollution as well as clouds. This is one reason why a satellite based approach is more appropriate for solar data sets vs. numerical weather prediction modeling, which is more appropriate for wind data.

How capable is the model in helping capture anomalous weather events, and using those to project weather patterns for future years?

Numerical weather prediction (NWP) models are very capable of capturing anomalous

events, both wind droughts and extremes such as cyclones. This is also true for the satellite based approaches, which were used to develop the solar data. Regarding “What does this mean for the future?”: One can look for trends in the last 10 years and then use experience to make an educated guess about the future, though care is needed if this is done because natural variability can suggest trends that subsequently reverse. How climate change might affect future weather patterns can be explored using global circulation models (GCMs) which are a different, but related, class of models to NWP. However, it should be noted that there is poor agreement between different GCMs about how wind and solar resources will evolve as greenhouse gas forcing raises Earth’s equilibrium temperature.

What model was used to split GHI into DHI and DNI?

The conversion from GHI to DNI and DHI is a proprietary Vaisala model. More details can be found in the methodology document.

What is the value of the shear exponent used to interpolate wind speed? Is it region specific or a constant for the entire country?

No shear exponent is used to interpolate wind — the profile is explicitly modeled, and changes from every pixel and every time step.

How might developers use these datasets for informing siting?

The weather datasets are a reasonable first approximation for figuring out any good resource areas that aren’t well documented (though most are at this point), and for assessing their climatological attributes. The weather data are a reasonable starting point, but the aggregated renewable datasets (described below) are too coarse for site selection use.

What is the uncertainty associated with these weather datasets?

Please refer to the brief methodology and validation document available at: gridlab.org/india_weatherdata/

Renewable datasets and power system modeling

Is it problematic that the aggregated processed RE datasets by NREL are broad geographically as compared to the raw weather datasets, which are at a 5 km resolution? Will this impact the results of the modeling, given the variability from location to location?

NREL’s approach to modeling RE generation is to develop a “representative profile” for each region. The representative profile is selected from all possible profiles at a 5 km resolution to best represent the average variability of resources within the region. It is worth noting that the representative profile is not an average profile, because this would mask the actual variability of the resource. Instead, the representative generation profile is based on a simulation of hourly operations for a utility-scale plant within the region (the method is described in detail below). This way, each region in the PCM “sees” a realistic generation profile based on the

prevalent weather variability in that region. Please note that this approach is not suitable for doing project development where a specific location will want to be compared to another, which would require on-the-ground measurements for additional accuracy.

How are the ‘representative’ renewable energy profiles calculated?

Representative profiles are selected for each ReEDS region by first calculating the generation for each 5 km weather node. The mean generation in the region is then calculated, weighting each weather node by the ratio of viable land that is available in its surrounding 5 km² cell. The generation profile of each point is then compared to the mean and the profile with the lowest least square difference from the mean profile is selected.

What can we do if we disagree with the NREL technology assumptions in developing the RE datasets?

You can adjust the processed datasets through “reverse” engineering as we have shared the technology assumptions. Alternatively, you can access the raw weather datasets and redo the technology analysis, for example, using the NREL System Advisor Model (SAM). There is a python PySAM package. More information here: sam.nrel.gov/.

Regarding wind turbine hub heights, the modeled data from Vaisala are at 28, 88, 161, 247, 351, and 500 meters (m). Downstream users of the data can interpolate the modeled values to calculate any hub height they want, and Vaisala has provided interpolated data at hub heights of 80m, 100m, 120m, and 140m for convenience. For the RE datasets, NREL interpolated data at a hub height of 135m for onshore wind and 150m for offshore wind.

Vaisala’s interpolation method was done as follows. For each 15 minute time step of the dataset, the profile between the bounding model levels (28m, 88m, 161m, 247m, 351m, 500m) of the variable of interest was calculated. That profile was used to calculate the value of that variable at that time step at the interpolated height. In the case of wind speed, the profile was assumed to follow a power law relationship, similar to what is described at the following link, with alpha calculated for each time step: en.wikipedia.org/wiki/Wind_profile_power_law

Is ReEDS India, which was referenced for doing capacity expansion modeling, an open-source model?

Yes it is an open source model available on GitHub (www.nrel.gov/analysis/reeds/). In addition, reV (Renewable Energy Potential model) is also open source www.nrel.gov/gis/renewable-energy-potential

Mechanics and data access themes

Can we access the processed RE datasets at the 5 km geography? If so, how?

We have recently made these datasets (labeled as “Granular Datasets”) available on the GridLab website under the same link shared earlier for the NREL processed data. Please note these are ~101 GB, about 25% larger than the samples of the raw weather datasets.

How can we visualize the raw or processed datasets?

The RE resource datasets are made available in csv format, and can be used for custom data visualization. We are also making available a QGIS file (.gqz format) that users can open up and toggle layers on and off to visualize the data overlaid on a map of India. QGIS is a free, open-source, and user-friendly Geographic Information System (GIS) software used for creating, editing, visualizing, analyzing, and publishing geospatial information.

The files are sorted by year. How can I access the raw data or processed data for a specific location for all weather years?

The meta data csv files for each technology contain information about which resource region corresponds to which state (and lists the latitude and longitude coordinates for the most “representative” data point in that subregion). This meta data applies to all years, and can be used to access data for a specific location for all weather years.

How can the raw weather datasets be accessed on Amazon Web Services?

We have shared the “URI” link on our website, which identifies the location of the raw datafiles on the AWS server (note, the URI is not the same as a URL). As we noted, we strongly recommend that you first test out the sample raw weather files, which are available on the website, before accessing files on the AWS server. Rather than aiming to download the full set of raw weather files, which are voluminous at 20 TB, and would take a very long time to download, it is advisable to explore computing in a cloud environment and/or high performance computing environment.

What should I do if I am new to AWS?

We suggest you follow these steps.

Step 1: You’ll need to register for an AWS account.

Step 2: You’ll need to install the [AWS CLI tool](#) and [set it up](#) using your AWS account. We recommend using “[short term credentials](#)”.

Step 3: You’ll need to run this command, filling in the file you want and where they want the file to go: `aws s3 cp s3://vaisala-india-data/{whatever file you want} {download destination} --request-payer`

More information on requester pays buckets is in [this link](#).

How can these datasets be used in different models?

The datasets are in universal format, and can be converted to whatever format your power system model requires.