

# Exploratory Data Analysis - Data Visualization

*Tenzin*

*10/3/2019*

## I. Where is the Data from?

The NSRDB 1991–2010 update is a serially complete collection of hourly values of the three most common measurements of solar radiation, i.e. Global Horizontal (GHI), Direct Normal (DNI), and Diffuse Horizontal Irradiance (DHI), over a period of time adequate to establish means and extremes and at a sufficient number of locations to represent regional solar radiation climates. Nearly all of the solar data in the NSRDB are modeled, and only 40 sites have measured solar data — none of them with a complete period of record.

The solar radiation that passes through directly to the earth's surface is called Direct Normal Irradiance (DNI). The radiation that has been scattered out of the direct beam is called Diffuse Horizontal Irradiance (DHI). The direct component of sunlight and the diffuse component of skylight falling together on a horizontal surface make up Global Horizontal Irradiance (GHI).

The SUNY model produces estimates of global and direct irradiance at hourly intervals on the 10-km grid for all 50 states, excluding Alaska above 60° north latitude and west of 160° west longitude, where the geostationary satellites cannot resolve cloud cover with necessary detail. SUNY model diffuse values were derived from the global and direct irradiance values for each grid cell as the difference between the global and vertical component of the direct. The directory structure that is divided into 2X2 degree tiles. Thus with 8-year files, there are 400 files in each subdirectory. For yearly files, there are 400 X 8 = 3200 files. So for example, directory 10640 has all files from -106 to -104 long, 40.0 to 42.0 lat. Within this will be a file such as `dir_glo_105454105.csv`, with 8 years of data from location 105.45 deg W, 41.05 degrees N. There are 436 of these subdirectories.

## II. What is time and spacial span?

The data we have here in the folder 726980 within Data, has the hourly solar radiation measurements from the Portland International Airport, OR from 1991 to 2005.

It was obtained from the National Solar Radiation Database (NSRDB) at the National Climatic Data Center (NCDC).

The manual for this data set can be obtained from the following extension: [https://www1.ncdc.noaa.gov/pub/data/nsrdb-solar/documentation-2010/NSRDB\\_UserManual\\_r20120906.pdf](https://www1.ncdc.noaa.gov/pub/data/nsrdb-solar/documentation-2010/NSRDB_UserManual_r20120906.pdf)

## III. What do the rows and columns stand for?

The columns represent the following Fields/Column Name/Units/Description:

1. Date/ YYYY-MM-DD/ Date of data record.
2. Time/ HH:MM/ Time of data record (local standard time).
3. Hourly mean zenith angle (for sunup periods)/ Decimal Degree/ Solar zenith angle (angle between sun and the zenith) as the mean of all 1-minute sunup zenith angle values for the 60-minute period ending at the timestamp.
4. Hourly mean azimuth angle (for sunup periods)/ Decimal degree/ Solar azimuth angle (angle between sun and north) as the mean of all 1-minute sunup azimuth angle values for the 60- minute period ending at the timestamp.
5. Hourly extraterrestrial radiation on a horizontal surface/ Watt-hour per square meter/ Amount of solar radiation received on a horizontal surface at the top of the atmosphere during the 60- minute period ending at the timestamp.

6. Hourly extraterrestrial radiation normal to the sun/ Watt-hour per square meter/ Amount of solar radiation received on a surface normal to the sun at the top of the atmosphere during the 60-minute period ending at the timestamp.
7. Modeled global horizontal/ Watt-hour per square meter/ Total amount of direct and diffuse solar radiation (modeled) received on a horizontal surface during the 60-minute period ending at the timestamp
8. Modeled global horizontal uncertainty/  $\pm\%$ / Section 1.4.5.1. National Solar Radiation Database 1991–2010 Update: User’s Manual
9. Modeled global horizontal source flag// Table 7. National Solar Radiation Database 1991–2010 Update: User’s Manual
10. Modeled direct normal/ Watt-hour per square meter/ Amount of solar radiation (modeled) received in a collimated beam on a surface normal to the sun during the 60-minute period ending at the timestamp
11. Modeled direct normal uncertainty/  $\pm\%$ / Section 1.4.5.1. National Solar Radiation Database 1991–2010 Update: User’s Manual
12. Modeled direct normal source flag// Table 8. National Solar Radiation Database 1991–2010 Update: User’s Manual
13. Modeled diffuse horizontal/ Watt-hour per square meter/ Amount of solar radiation (modeled) received from the sky (excluding the solar disk) on a horizontal surface during the 60- minute period ending at the timestamp
14. Modeled diffuse horizontal uncertainty/  $\pm\%$ / Section 1.4.5.1. National Solar Radiation Database 1991–2010 Update: User’s Manual
15. Modeled diffuse horizontal source flag// Table 8. National Solar Radiation Database 1991–2010 Update: User’s Manual
16. Measured global horizontal/ Watt-hour per square meter/ Total amount of direct and diffuse solar radiation received on a horizontal surface during the 60- minute period ending at the timestamp
17. Measured global horizontal quality flag/ SERI-QC/ Table 11. National Solar Radiation Database 1991–2010 Update: User’s Manual
18. Measured direct normal/ Watt-hour per square meter/ Amount of solar radiation received in a collimated beam on a surface normal to the sun during the 60- minute period ending at the timestamp
19. Measured direct normal quality flag/ SERI-QC/ Table 11. National Solar Radiation Database 1991–2010 Update: User’s Manual
20. Measured diffuse horizontal Unit/ Watt-hour per square meter/ Amount of solar radiation received from the sky (excluding the solar disk) on a horizontal surface during the 60-minute period ending at the timestamp
21. Measured diffuse horizontal quality flag/ SERI-QC/ Table 11. National Solar Radiation Database 1991–2010 Update: User’s Manual
22. \*Total sky cover/ Tenth of sky/ Amount of sky dome covered by clouds or obscuring phenomena at the time indicated. Section 1.4.5.2. National Solar Radiation Database 1991–2010 Update: User’s Manual
23. \*Precipitable water/ Centimeter/ The total precipitable water contained in a column of unit cross section extending all of the way from the earth’s surface to the “top” of the atmosphere. Section 2.2.1.4. National Solar Radiation Database 1991–2010 Update: User’s Manual
24. \*Precipitable water flag// Section 1.4.5.2. and Table 13. National Solar Radiation Database 1991–2010 Update: User’s Manual

25. \*Aerosol optical depth, broadband/ Unitless/ The broadband optical depth per unit of air mass due to extinction by the aerosol component of the atmosphere (randomly perturbed). Section 2.2.1.2.National Solar Radiation Database 1991–2010 Update: User’s Manual
26. \*Aerosol optical depth flag// Section 2.2.1.2.National Solar Radiation Database 1991–2010 Update: User’s Manual

#### IV. Preliminary visualizations (Time vs. Radiation).

```
library(readr)
library(tidyverse)

## -- Attaching packages -----
## v ggplot2 3.2.1      v purrr  0.3.2
## v tibble  2.1.1      v dplyr  0.8.0.1
## v tidyr   0.8.3      v stringr 1.4.0
## v ggplot2 3.2.1      v forcats 0.4.0

## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library("dplyr")

X726980_1991 <- read_csv('/home/tensangpo/reed-thesis/Data/726980/726980_1991.csv')

## Parsed with column specification:
## cols(
##   .default = col_double(),
##   `YYYY-MM-DD` = col_date(format = ""),
##   `HH:MM (LST)` = col_time(format = "")
## )

## See spec(...) for full column specifications.

## Warning: 365 parsing failures.
## row      col      expected actual                                     file
## 24 HH:MM (LST) valid date  24:00 '/home/tensangpo/reed-thesis/Data/726980/726980_1991.csv'
## 48 HH:MM (LST) valid date  24:00 '/home/tensangpo/reed-thesis/Data/726980/726980_1991.csv'
## 72 HH:MM (LST) valid date  24:00 '/home/tensangpo/reed-thesis/Data/726980/726980_1991.csv'
## 96 HH:MM (LST) valid date  24:00 '/home/tensangpo/reed-thesis/Data/726980/726980_1991.csv'
## 120 HH:MM (LST) valid date  24:00 '/home/tensangpo/reed-thesis/Data/726980/726980_1991.csv'
## ... ..
## See problems(...) for more details.

X726980_1991 <- X726980_1991 %>% mutate(Time = row_number())

# View(X726980_1991)
```

The intent of the modeled data is to present hourly solar radiation values that, in the aggregate, possess statistical properties (e.g., means, standard deviations, and cumulative frequency distributions) that are as close as possible to the statistical properties of measured solar data over the period of a month or year. These data do not represent each specific hourly value of solar radiation to the same or equivalent accuracy as the long-term statistics.

```
Y <- 'Dir Mod (W/m^2)'

ggplot(data = X726980_1991, mapping = aes(x = Time, y = Y)) + geom_line() + geom_point() + xlab("Time (l
```

