datascience+ **Time Series Analysis with Wind Resource...**

onlinecourses . printedbooks

An online community for showcasing R & Python tutorials

- About Us
- Archives
- Contribute
- Sign Up
- Log In
- [Search] Search

- Advanced Modeling in R

Time Series Analysis with Wind Resource Assessment in R

- Published on August 2, 2019 at 10:58 pm
-

 Introduction Getting Data Data Management Visualizing Data Basic Statistics Regression Models Advanced Modeling Programming Tips & Tricks Video Tutorials

One of the sectors with a huge demand for data science/analysis is the energy sector. A branch of this sector where demand is high is the green wind energy turbine sector. In this analysis, you will learn to do Time Series Analysis with Wind Ressource Assement in R.

Wind Ressource Assement

The wind is a highly variable resource. In the time domain, there are gusts acting over seconds, calm periods lasting for days, seasonal patterns across the year and climatic influences leading to interannual variation. In the spatial domain, trees cause very localized shelter effects, gentle hills can funnel the wind smoothly around them and continental landmasses slow down the winds that whip across the ocean. Consequently, assessing the wind resource is not easy.

The power available in the wind is derived from the kinetic energy available in a stream tube passing through a given area. The most important part of wind resource assessment is the fact that the available power is proportional to the cube of the wind speed:

$$P = ½ * \rho A V^3$$

Where:
P = power available (W)
ρ = density of air (kg/m3: approx. 1.2kg/m3 at sea level &
15°C)
A = swept area of wind energy conversion device (m2: = πr2 for
conventional horizontal axis rotors)
V = wind velocity (m/s)

This cubic relationship is the single most important point relating to the assessment of the wind resource, as a doubling of the wind speed yields an eight-fold increase in power! As a result, an accurate assessment of the wind resource at each proposed site is absolutely vital (much more so than for solar, where the power produced is linearly proportional to the resource), as even, a relatively small difference in the resource can make or break a project.

Time Series Data

A time series is a collection of observations of well-defined data items obtained through repeated measurements over time.

For example, measuring the level of wind per m/s by the daily, monthly time of the year would comprise a time series. This is because wind per m/s are well defined and consistently measured at equally spaced intervals. Data collected irregularly or only once are not time series.

A times series allows you to identify change within a population over time. A time series can also show the impact of cyclical, seasonal and irregular events on the data item being measured.

It is very important for the wind industry to be able to describe the variation of wind speeds. Turbine designers need the information to optimize the design of their turbines, so as to minimize generating costs. Turbine investors need the information to estimate their income from electricity generation.

If you measure wind speeds throughout a year, you will notice that in most areas strong gale force winds are rare, while moderate and fresh winds are quite common.
The wind variation for a typical site is usually described using the Weibull distribution. A wind site has a mean wind speed of a specific meter per second, and the shape of the curve is determined by a shape parameter.

Wind Ressource Assesment in R

The goal here is to illustrate how aspects of typical wind resource assessment and energy capture from meteorological data can be accomplished using open source tools, in this case using R. Using publicly available data, we will walk through some of the typical steps taken in site screening, importing, visualizing and analyzing meteorological data with the goal of modeling the annual energy capture of a wind turbine at a given location.

First, let us load the R packages used to do the analysis.

```
library(tidyverse) # r-package ecosystem
library(MASS) # statistical functions
library(knitr) # fancy tables
library(clifro) # windrose plot
library(scales) # percentages
library(elevatr) # elevation data
library(raster) # geospatial
library(leaflet) # mapping
Copy
```

Get Wind Data in R

Next, we'll download some sample wind resource data from NREL's wind toolkit to see how energetic my location is. The variable arguments in the paste statement below should be replaced with your own information (see https://developer.nrel.gov/signup/ to get an API key).

```
lat=42.2756887
lon=-71.21161789999996
url<-paste('http://developer.nrel.gov/api/wind-toolkit/wind/wtk_download.csv?api_key=',api_key,'&wkt=POINT(',lon,'%20',lat,
')&attributes=wind_speed,wind_direction,power,temperature,pressure&names=2009&full_name=',name,'&email=',email,'&affiliation=',affiliation,
'&reason=Example',sep='')

#download data as a dataframe
df<-read_csv(url,skip=3,col_types = cols())
# tidy names
names(df)<-gsub("[^[:alnum:]]", "", names(df))
# convert dates to timestamp
df$timestamp_utc=as.POSIXct(paste(df$Year,df$Month,df$Day,df$Hour,df$Minute,'00',sep='-'),format='%Y-%m-%d-%H-%M-%S',tz='UTC')
Copy
```

Exploratory Analysis in R

Now to the meat of the post, which is to illustrate how to analyze and visualize typical steps for wind resource assessment.

```
#Data Structure
head(df)
## # A tibble: 6 x 11
##    Year Month  Day  Hour Minute windspeedat100m… winddirectionat… powerMW
##
## 1  2009     1    1     0      0           13.3              318    16.0
## 2  2009     1    1     0      5           13.3              317    16.0
## 3  2009     1    1     0     10           13.4              317    16.0
## 4  2009     1    1     0     15           13.4              317    16.0
## 5  2009     1    1     0     20           13.5              317    16.0
## 6  2009     1    1     0     25           13.6              317    16.0
## # ... with 3 more variables: airtemperatureat2mK ,
## #   surfaceairpressurePa , timestamp_utc
Copy
```

Now it is time to make the time series analysis in R:

```
ggplot(df,aes(timestamp_utc,windspeedat100mms))+geom_line()+theme_minimal()
Copy
```

The above coding gives us the following graph:

Monthly Wind Speed Distribution in R

```
ggplot(df,aes(factor(Month),windspeedat100mms))+
  geom_boxplot()+
  theme_minimal()+
  labs(x='Month')
Copy
```

The above coding gives us the following graph:

Wind Rose in R

A wind rose is a meteorological diagram depicting the distribution of wind direction and speed at a location over a period of time. It is possible to make a wind rose in R with the following code:

```
windrose(speed = df$windspeedat100mms,
              direction = df$winddirectionat100mdeg,
              n_directions = 12,
              speed_cuts = seq(0,20,4),
              ggtheme='minimal',
              col_pal = 'YlGnBu')
Copy
```

Weibull Distrubution for Wind Data in R

It is very important for the wind industry to be able to describe the variation of wind speeds. Turbine designers need the information to optimise the design of their turbines, so as to minimise generating costs. Turbine investors need the information to estimate their income from electricity generation. Consider the following graph:

The above statistical graph shows a probability density distribution. The area under the curve is always exactly 1 since the probability that the wind will be blowing at some wind speed including zero must be 100 percent.
Half of the blue area is to the left of the vertical black line at 8 meters per second. The 8 m/s is called the median of the distribution. This means that half the time it will be blowing less than 8 meters per second, the other half it will be blowing faster than 8 meters per second.
You may wonder then, why we say that the mean wind speed is 8 meters per second. The mean wind speed is actually the average of the wind speed observations.
As you can see, the distribution of wind speeds is skewed, i.e. it is not symmetrical. Sometimes you will have very high wind speeds, but they are very rare. Wind speeds of 9 meters per second, on the other hand, are the most common ones. 9 meters is called the modal value of the distribution. If we multiply each tiny wind speed interval by the probability of getting that particular wind speed and add it all up, we get the mean wind speed.
The statistical distribution of wind speeds vary from place to place around the globe, depending upon local climate conditions, the landscape, and its surface. The Weibull distribution may thus vary, both in its shape and in its mean value.

Another way of finding the mean wind speed is to balance the Weibull Distribution to the right, which shows exactly the same as the graph above. Each brick represents the probability that the wind will be blowing at that speed during 1 percent of the time during the year. 1 m/s wind speeds are in the pile to the far left, 26 m/s is to the far right.
The point at which the whole pile will balance exactly will be at the 8th pile, i.e. the mean wind speed is 8 m/s.

In order to fit a Weibull Distribution in R use the following code:

```
weibull_fit<-fitdistr(df$windspeedat100mms,'weibull')
x<-seq(0,20,.01)
weibull_density<-tibble(x,y=dweibull(x = x,shape = weibull_fit$estimate[1],scale = weibull_fit$estimate[2]))
ggplot(df,aes(windspeedat100mms))+
  geom_histogram(aes(y=..density..),bins=30,color='white')+
  geom_line(data=weibull_density,aes(x=x,y=y),color='red')+
  theme_minimal()
Copy
```

datascience+ **Time Series Analysis with Wind Resource...**   Share: Twitter
                                                                        Facebook

Energy Capture in R

Renewable energy (sources) or RES capture their energy from existing flows of energy, from on-going natural processes, such as sunshine, wind, flowing water, biological processes, and geothermal heat flows.

The most common definition is that renewable energy is from an energy resource that is replaced rapidly by a natural process such as power generated from the sun or from the wind.

Most renewable forms of energy, other than geothermal and tidal power, ultimately come from the Sun.

Some forms are stored solar energy such as rainfall and wind power which are considered short-term solar-energy storage, whereas the energy in biomass is accumulated over a period of months, as in straw, or through many years as in wood.

Capturing renewable energy by plants, animals, and humans do not permanently deplete the resource.

Let's model energy capture for this location in R:

Power Curve in R

Let's use the GE 1.5SLE 77m turbine power curve as an example:

```
url<-'http://www.wind-power-program.com/Downloads/Databasepowercurves(May2017).zip'
tmp<-tempfile()
download.file(url,tmp)
unzip(tmp,files = 'Databasepowercurves(May2017)/HAWTs/500kw and above/General Electric/GE 1.5SLE 77m 1.5MW (MG).pow',junkpaths = T)
unlink(tmp)
pc<-read_csv('GE 1.5SLE 77m 1.5MW (MG).pow',
           skip=5,col_names = F,col_types='n',n_max = 30)
pc<-tibble(ws=seq(0.5,29.5,1),kw=pc$X1)
ggplot(pc,aes(ws,kw))+geom_point()+geom_line()+theme_minimal()
Copy
```

The above coding gives us the following graph:

datascience+ **Time Series Analysis with Wind Resource...**

datascience+Time Series Analysis with Wind Resource...

It's important to adjust the raw wind speed by air density as wind power density is a function of air density.

```
df<-mutate(df,air_density=(df$surfaceairpressurePa/df$airtemperatureat2mK*287)*.00001,
           dc_ws=windspeedat100mms*(air_density/mean(air_density))^(1/3))
ggplot(df,aes(windspeedat100mms,dc_ws,color=air_density))+
  geom_point()+
  theme_minimal()+
  coord_equal()
Copy
```

The above coding gives us the following graph:

Predict Energy in R

We'll use R's approx function to interpolate the equivalent turbine power for each wind speed in the time series. There's lot of different methods for this of course.

```
df$kw<-approx(pc$ws,pc$kw,df$dc_ws)$y
ggplot(df,aes(kw))+
  geom_density(fill='blue')+
  theme_minimal()
Copy
```

The above coding gives us the following graph:

Results in R

Now let us present the results in our model in R:

```
#Aggregates
monthly%
  group_by(Month) %>%
  summarise(hours=n()/12,
            windspeedat100mms=mean(windspeedat100mms),
```

Share: Twitter
                                                                                        Facebook

```
Month    hours   windspeedat100mms       mwh     ncf
1        744     6.8     467.9   41.9%
2        672     8.3     584.4   58.0%
3        744     7.2     510.5   45.7%
4        720     7.4     512.2   47.4%
5        744     6.7     420.7   37.7%
6        720     5.2     239.8   22.2%
7        744     6.2     354.7   31.8%
8        744     5.5     285.9   25.6%
9        720     6.4     396.3   36.7%
10       744     7.3     504.5   45.2%
11       720     7.1     471.7   43.7%
12       744     8.7     691.6   62.0%
Copy
```

```
#annual
annual<-data.frame(mwh=sum(monthly$mwh),ncf=percent(sum(monthly$mwh/(sum(monthly$hours)*1.5))))
kable(annual,digits=1,caption = 'annual totals',align='c')
mwh     ncf
5440.1  41.4%
Copy
```

Nice, a 41.4% NCF for this location is fine.

Comparison with NREL Model in R

NREL's dataset comes with an example of power measurement based on a 5MW wind turbine. Let's see how our R model stacks up using the GE 1.5SLE.

```
ggplot(df,aes(kw*.001,powerMW))+
  geom_point()+
  theme_minimal()+
  labs(x='GE 1.5sle MW',y='NREL 5MW turbine example MW')
Copy
```

The above coding gives us the following graph:

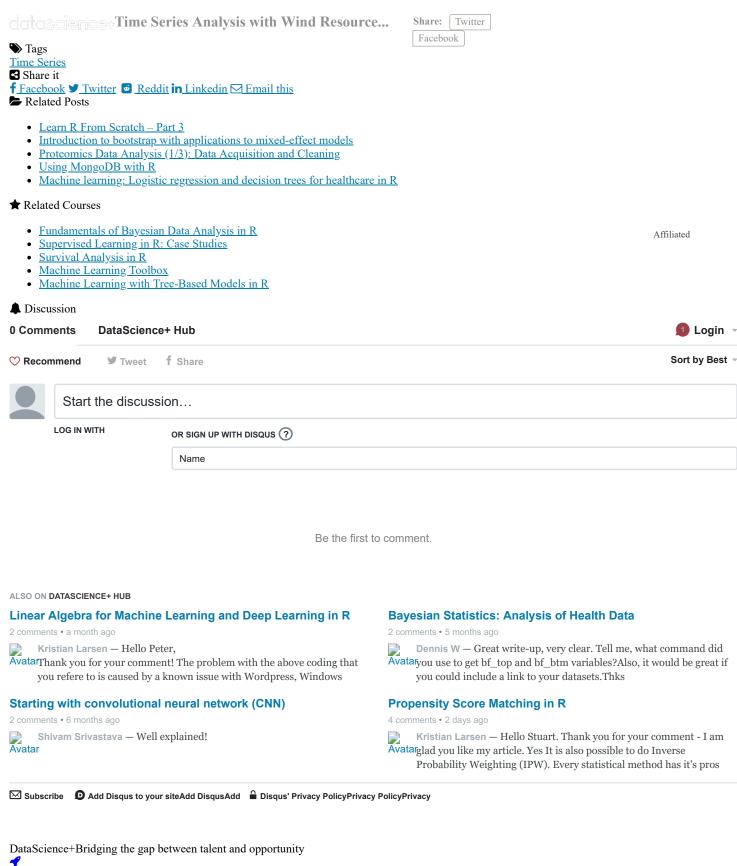As we see in the above graph our model in R is close to the model of NREL.

Author

Kristian Larsen
Economic data scientist

More from Author

- Program Evaluation: Regression Discontinuity Design in R
- Propensity Score Matching in R
- Multilevel Models in R

Disclosure

datascience-**Time Series Analysis with Wind Resource...**

🏷 **Tags**

Time Series

⊠ **Share it**

f Facebook **🐦** Twitter **🤖** Reddit **in** Linkedin **✉** Email this

📂 **Related Posts**

- Learn R From Scratch – Part 3
- Introduction to bootstrap with applications to mixed-effect models
- Proteomics Data Analysis (1/3): Data Acquisition and Cleaning
- Using MongoDB with R
- Machine learning: Logistic regression and decision trees for healthcare in R

⭐ **Related Courses**

- Fundamentals of Bayesian Data Analysis in R
- Supervised Learning in R: Case Studies
- Survival Analysis in R
- Machine Learning Toolbox
- Machine Learning with Tree-Based Models in R

Affiliated

🔔 **Discussion**

0 Comments	DataScience+ Hub	① Login

♡ Recommend   🐦 Tweet   f Share   Sort by Best

Start the discussion…

LOG IN WITH   OR SIGN UP WITH DISQUS ⑦

Name

Be the first to comment.

Linear Algebra for Machine Learning and Deep Learning in R

2 comments • a month ago

Kristian Larsen — Hello Peter,
Avatar Thank you for your comment! The problem with the above coding that you refere to is caused by a known issue with Wordpress, Windows

Bayesian Statistics: Analysis of Health Data

2 comments • 5 months ago

Dennis W — Great write-up, very clear. Tell me, what command did
Avatar you use to get bf_top and bf_btm variables?Also, it would be great if you could include a link to your datasets.Thks

Starting with convolutional neural network (CNN)

2 comments • 6 months ago

Shivam Srivastava — Well explained!
Avatar

Propensity Score Matching in R

4 comments • 2 days ago

Kristian Larsen — Hello Stuart. Thank you for your comment - I am
Avatar glad you like my article. Yes It is also possible to do Inverse Probability Weighting (IPW). Every statistical method has it's pros

DataScience+Bridging the gap between talent and opportunity
🚀

>_ **Site Links**

- About Us
- Archives
- Contribute
- R Markdown

>_ **Legal**

datascience+ **Time Series Analysis with Wind Resource...**

- [Contact Us](#)

>_ Articles

- [Introduction](#)
- [Getting Data](#)
- [Data Management](#)
- [Visualizing Data](#)
- [Basic Statistics](#)
- [Regression Models](#)
- [Advanced Modeling](#)
- [Programming](#)

- Facebook
- Twitter
- LinkedIn
- RSS

Connect with Us