TEACHING STATISTICAL METHODS IN ENGINEERING COURSES THROUGH WIND POWER DATA

Jose Francisco Moreira Pessanha,[1] **Fernando Luiz Cyrino Oliveira,**[2] **Reinaldo Castro Souza**[3]

ABSTRACT

We describe examples based on real wind speed data designed to introduce engineering students at the post-calculus level to statistical methods and theory with real engineering problems. The examples cover some steps of the traditional wind power data analysis in order to develop in the student the data analysis capabilities and the statistical reasoning applied to the engineering problems.

Keywords: Teaching Statistics; R project; Wind Power Data.

RESUMO

Neste artigo, descrevemos exemplos baseados em dados de velocidade real do vento, apropriados para introduzir os estudantes de engenharia, em nível de pós-cálculo, aos métodos e à teoria estatísticos, a partir de problemas reais de engenharia. Os exemplos cobrem algumas etapas da análise tradicional de dados de energia eólica, a fim de desenvolver no aluno a capacidade de análise de dados e do raciocínio estatístico aplicado aos problemas de engenharia.

Palavras-chave: Ensino de Estatística; Projeto R; dados de energia eólica.

INTRODUCTION

In general, the engineering students take only one semester course about Probability and Statistics, a short period to learn and develop the statistical reasoning applied to the engineering problems. In this context, it is rather important to motivate students with examples from real problems found in the engineering practice in order to show that statistics play an important role in the engineering.

This paper aims to describe a set of activities based on real wind speed data from public sources and designed to provide hands-on exercises in the classroom with R project (R Core Team, 2014). Naturally the examples presented in this paper can be tailored to other computational languages. The examples are oriented to introduce students to statistical concepts and methods in special, data exploratory analysis, descriptive statistics and statistical inference. The main outcome that we expect to achieve with the adoption of these examples is to motivate and develop the data analysis capabilities and statistical reasoning of the students.

In addition, the proposed examples have important practical implications: it can develop and enhance the programming skills of the students, improve their abilities to solve problems as well as to point out links with other disciplines like electromechanical energy conversion, electric machines, power generation, power system control & operation and power system economics. This initiative is an original idea that offers new ways to teach sta-

---

1 D.Sc. in Electrical Engineering, Adjunct Professor, State University of Rio de Janeiro (UERJ); professorjfmp@hotmail.com

2 D.Sc. in Electrical Engineering, Assistant Professor, Industrial Engineering Department, Pontical Catholic University of Rio de Janeiro (PUC-Rio); cyrino@puc-rio.br

3 Ph.D. in Statistics, Full Professor, Electrical Engeneering Department, Pontical Catholic University of Rio de Janeiro (PUC-Rio); reinaldo@ele.puc-rio.br

tistics to engineering students and show them the power of statistics at same time that present wind power, a valuable energetic resource.

WIND POWER

The statistics from World Wind Energy Association (WWEA, 2014) show that the total worldwide installed wind capacity reached 336 GW by mid-2014, a growth of 13.5% over mid-2013. Wind is a renewable natural resource and its utilization in the electricity production is one of the most promising alternatives to mitigate air pollution in the electric power generation.

However the wind is a non-controllable resource, it has a stochastic behavior and to cope with its variability the engineers apply statistical and probabilistic methods in the design of wind turbines and management of wind power plants.

The relationship between the wind speed v (m/s) and the power generation P (Watts) is defined by the power curve $P(v)$ illustrated in Figure 1. Examples of real power curves can be obtained in the catalogs of manufacturers of wind turbines available on the internet. The power generation begins at cut-in speed typically between 2 and 3m/s. Thus, the power output rises rapidly up to achieve the rated output power at the rated wind speed between 12 and 17m/s. At higher wind speeds the turbine holds the power generation near to the rated output, but when the turbine reach the cut-out wind speed around 25m/s the wind turbine power generation must be turned off in order to avoid risk of rotor damage. For assessment of wind power potential of a site we need to know the power curve parameters of the wind turbine (cut-in, rated and cut-out speeds) and the parameters of the wind speed distribution (usually Weibull or Rayleigh) estimated from on-site wind speed measurements at hub-height (1).

Figure 1 – Theoretical power curve.
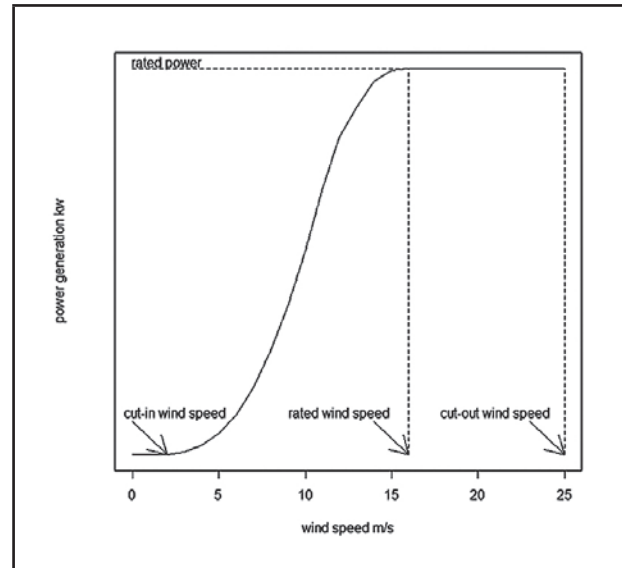
$$P(v) = \begin{cases} 0, v \leq V_{in} \\ 0.5 C_p \cdot \eta_m \cdot \eta_g \cdot \rho \cdot A \cdot v^3, V_{in} < v \leq V_r \\ 0.5 C_p \cdot \eta_m \cdot \eta_g \cdot \rho \cdot A \cdot V_r^3, V_r < v \leq V_{out} \\ 0, v > V_{out} \end{cases} \quad (1)$$

where:
V_{in} = cut-in wind speed (m/s)
V_r = rated wind speed (m/s)
V_{out} = cut-out wind speed (m/s)
C_p = coefficient of performance (dimensionless)
η_m = generator efficiency (dimensionless)
η_g = gearbox efficiency (dimensionless)
ρ = air density (kg/m³)
A = rotor swept area (m²)

Source: JANGAMSHETTI & RAU, 1999.

EXAMPLES

Below we present descriptions of four activities based on real wind speed data from public sources and designed to provide hands-on exercises in the classroom with R project.

Activity 1: Download wind speed data from public sources

There are several links on internet where we can find wind speed measurements available. For example, in Brazil the National Network of Environmental Data for Renewable Energy Resource Assessment (SONDA project) provides wind speed measurements from some Brazilian anemometric stations with ten minutes time resolution. The data are arranged on monthly electronic spreadsheets that can be downloaded free. This is a great opportunity to students learn how download data from internet through R. The following the code snippets downloads 12 monthly electronic spreadsheets with wind speed measurements from anemometric station located in São João do Cariri (7º22'54"S , 36º31'38"W), Brazil.

```
library(downloader) ; library(utils) # load packages
setwd("c:/IASE") # set work directory
year = 2006 ; station="SCR"
address = paste("http://sonda.ccst.inpe.br/basedados/dados/eolicos/SCR/",year,"/",sep="")
windspeed25 = numeric(0) ; windspeed50 = numeric(0) ; windmonth = numeric(0) # wind speed at 25m and 50m
height
for (month in 1:12) {
  if (month > 9) file = paste(station,"0",year-2000,month,"WD",sep="") # namefile
  if (month < 10) file = paste(station,"0",year-2000,"0",month,"WD",sep="") # namefile
  url = paste(address,paste(file,".zip",sep=""),sep="") # address
  download(url,file,mode="wb") # download file
  # extract csv and read csv file
  unzip(file,files=NULL,list=FALSE,overwrite=TRUE,junkpaths=FALSE,exdir="",unzip="internal",setTimes=FALSE)
  data = read.csv(paste(file,".csv",sep=""),sep=";",header=FALSE)
  windmonth = c(windmonth,rep(month,dim(data)[1])) # month
  windspeed25 = c(windspeed25,data[,6]) ; windspeed50 = c(windspeed50,data[,9]) # speed at 25m and 50m
  heigth
}
windspeed25[which(windspeed25<0)] = 0 ; windspeed50[which(windspeed50<0)] = 0 # replace negative speeds
by zero
```
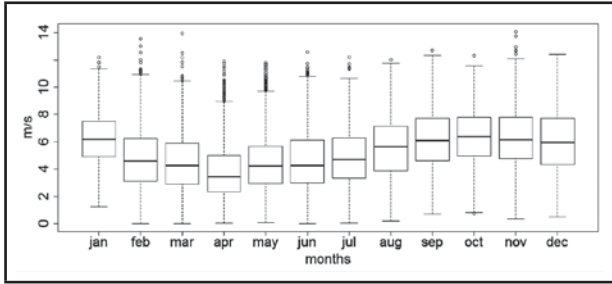
Activity 2: Exploratory data analysis

Based on the wind speed measurements the students can make an exploratory data analysis and compute some descriptive statistics from the wind speed dataset. For example, they can use the following R code to make the boxplots across the 12 months in Figure 2 and the descriptive statistics by month in Figure 3.

```
boxplot(windspeed50~ windmonth,xlab="months",ylab="m/s",xaxt= "n",cex.axis=2,cex.lab=2)
axis(1,at=1:12,labels=c("jan","feb","mar","apr","may","jun","jul","aug","sep","oct","nov","dec"),cex.axis=2)
aggregate(windspeed50,by=list(windmonth),FUN=summary)
```

Figure 2 – Boxplots of wind speed at 50 m height.

Source: R Core Team, 2014.

Figure 3 – Descriptive statistics.

	Group.1	x.Min.	x.1st Qu.	x.Median	x.Mean	x.3rd Qu.	x.Max.
1	1	1.240	4.918	6.200	6.239	7.502	12.170
2	2	0.000	3.098	4.600	4.816	6.240	13.560
3	3	0.000	2.880	4.260	4.480	5.903	13.930
4	4	0.040	2.350	3.430	3.770	4.990	11.880
5	5	0.050	2.960	4.240	4.403	5.660	11.810
6	6	0.000	2.988	4.260	4.654	6.130	12.570
7	7	0.020	3.320	4.710	4.795	6.282	12.200
8	8	0.190	3.890	5.630	5.555	7.132	12.010
9	9	0.690	4.630	6.100	6.157	7.730	12.710
10	10	0.750	4.978	6.380	6.386	7.780	12.310
11	11	0.360	4.780	6.170	6.336	7.800	14.080
12	12	0.480	4.330	5.970	6.072	7.730	12.420

Source: R Core Team, 2014.

Figures 2 and 3 show that the lowest median and mean wind speeds are observed in April and May. In addition, the average wind speed increases during the dry season (between May and November), but it follows a downward trend in the wet season (between December and April). This is an example of the energy complementarity between wind and hydro sources in Brazil (PALFI & ZAMBON, 2013).

Figure 4 shows monthly histograms for the wind speed at 50m height. The histograms were generated by the following R code. The distribution shape changes across the months and the statistics in Figure 5 indicate moderate skewed distribution in special at April and May when the kurtosis coefficient indicates moderate leptokurtic distribution.

```
par(mfrow=c(2,6)) ; labels=c("jan","feb","mar","apr","may","jun","jul","aug","sep","oct","nov","dec")
for (i in 1:12) {
    m=which(windmonth==i)
    hist(windspeed50[m],main=labels[i],xlab="m/s",cex.axis=1,cex.lab=1,cex.main=2)
}
```

Figure 4 – Monthly histograms.
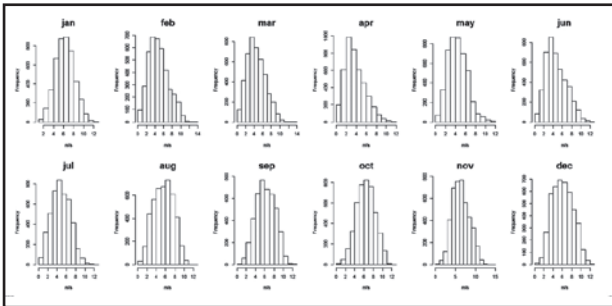
Source: R Core Team, 2014.

Figure 5 – Monthly standard deviation, skewness and kurtosis.

|	Standard Deviation	Skewness	Kurtosis
[1,]	1.851128	0.113531272	2.652212
[2,]	2.286779	0.440525105	2.680806
[3,]	2.125612	0.441204502	2.789647
[4,]	1.980068	0.742702546	3.300896
[5,]	1.900126	0.496970306	3.211324
[6,]	2.197749	0.514656235	2.668262
[7,]	2.003935	0.151759990	2.557935
[8,]	2.112200	-0.010098854	2.254257
[9,]	2.052793	0.005248375	2.385994
[10,]	1.986095	-0.036983955	2.609298
[11,]	2.095818	0.262305332	2.571293
[12,]	2.216109	0.159274157	2.323115

Source: R Core Team, 2014.

```
library(moments) # load package moments
others_stats=matrix(0,12,3)
colnames(others_stats)=c("Standard Deviation","Skewness","Kurtosis")
for (i in 1:12) {
 m=which(windmonth==i)
 ws = windspeed50[m]
 others_stats[i,]=c(sd(ws),skewness(ws),kurtosis(ws))
}
print(others_stats)
```

Activity 3: Probability distribution fitting

The Weibull probability distribution is the standard probabilistic model to describe the wind speed (JUSTUS *et al.*, 1978). The results from exploratory analysis indicate that the wind speed data are moderately skewed. Then, in this case the normal distribution is not the more appropriated model. The Weibull density function has two parameters as indicated in Figure 6: the shape parameter (k) and the scale parameter (c). The scale parameter (c) is proportional to the wind speed average (2) while the shape parameter (k) is related to the constancy of the wind, a higher k value imply in a lower wind speed variance (3). Additionally, the scale parameter (c) is expressed in m/s and the shape parameter (k) is dimensionless. In general k is a number between 2 and 3, but in the areas where prevailing trade winds like in the Brazilian Northeast the shape parameter can be greater than 6 (AMARANTE *et al.*, 2001).

Figure 6 – Weibull probability distribution model.
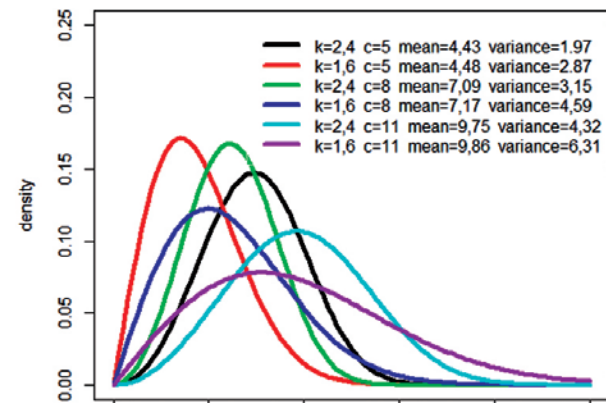
$$f(v) = \frac{k}{c}\left(\frac{v}{c}\right)^{k-1} \exp\left[\left(-\frac{v}{c}\right)^{k}\right] \quad v > 0 \quad (2)$$

$$E(v) = c\Gamma\left(1 + \frac{1}{k}\right) \quad (3)$$

where
$v \sim \text{Weibull}(k,c)$

The parameters c and k can be estimated by maximum likelihood (*parweibull* function from MASS package) or by the method of moments. In this paper we show the approach proposed by Blischke & Scheur (1986) to the method of moments where the parameters are determined in order to satisfies the following equations where $\overline{X}$ and S^2 are the wind speed sample statistics:

$$\overline{X} = c\Gamma\left(1 + \frac{1}{k}\right) \quad (4)$$

$$S^2 = c^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma\left(1 + \frac{1}{k}\right)^2\right] \quad (5)$$

The shape parameter estimate $\hat{k}$ is the root of the transcendental equation (6) and the scale parameter estimate $\hat{c}$ is obtained from equation (2) replacing k by $\hat{k}$ as shown in equation (7). Figure 7 shows the monthly estimates for k and c obtained by the R code below.

$$\frac{S^2}{\overline{X}^2} = \left(\Gamma\left(1 + \frac{2}{\hat{k}}\right)\middle/\Gamma\left(1 + \frac{1}{\hat{k}}\right)^2\right) - 1 \quad (6)$$

$$\hat{c} = \overline{X}\middle/\Gamma\left(1 + \frac{1}{\hat{k}}\right) \quad (7)$$

Figure 7 – Monthly estimates for the shape and scale parameters of the wind speed at 50m height.

```
Blischke_Scheur <- function(k){
   return(gamma(1+2/k)/gamma(1+1/k)^2 - 1 - ws_var/ws_mean^2)
}
k=rep(0,12);c=rep(0,12)
for (i in 1:12) {
   m=which(windmonth==i) ; ws = windspeed50[m]
   ws_mean=mean(ws) ; ws_var=var(ws)
   k[i] = uniroot(Blischke_Scheur,c(1,10))$root
   c[i] = ws_mean/gamma(1+1/k[i])
}
        cbind(k,c)
```

Activity 4: Estimating the capacity factor

There are many manufacturers and types of wind turbines and an engineer should choice the turbine model more appropriated to the wind behavior in a wind power plant project. A key parameter in the choice of turbine model is the capacity factor (CF) defined by the ratio between the average power output $E(P(v))$ and the rated power output P_{max}. Jangamshetti & Rau (1999) presents the following formula (8) to estimate the capacity factor of a wind turbine:

$$CF = \frac{E(P(v))}{P_{max}} = \frac{1}{v_r^3}\int_{V_{in}}^{V_r} v^3 f(v)dv + \int_{V_r}^{V_{out}} f(v)dv \qquad (8)$$

where $f(v)$ is the probability density function of the wind speed at hub height and V_{in}, V_r and V_{out} are respectively the cut-in, rated and cut-out speeds in the power curve (as shown in Figure 1).

In the modern wind power plants the hub height of the turbines can achieve more than 100m height, but the wind speed measurements from SONDA are taken at 25m and 50m heights. Then, it is necessary to extrapolate the estimates for the Weibull parameters to the hub height. This is an example of function of a random variable. The relationship between the wind speed and height can be modeled by the power law (JANGAMSHETTI & RAU, 1999) in which the wind speed ratio at two different heights (h and 50 m) is a power function of the ratio between the respective heights (9):

$$\frac{V_h}{V_{50m}} = \left(\frac{h}{50}\right)^{\alpha} \qquad (9)$$

$$\hat{\alpha} = \frac{\log\left(\overline{V_{50m}}\right) - \log\left(\overline{V_{25m}}\right)}{\log(50) - \log(25)} \qquad (10)$$

where the power law index α can be determined by the equation (10) based on mean wind speed at 25 m and 50 m height.

Based on power law we can extrapolate the wind speed measurements to a hub height h and then apply the method of moments to estimate the parameters k and c at new height h. Instead, we can apply the following corrections presented by Justus et al. (1978).

Consider a wind turbine with rated power equal to 1.8 MW, hub height 119m and which cut-in speed, rated speed and cut-out speed are 3m/s, 12m/s and 25m/s respectively. In order to calculate the capacity factor we should correct the parameters of the wind speed probability density function from 50 m height to 119 m height. The following R code applies the equations (11) and (12) to correct the parameters k and c and then compute the monthly capacity factor.

$$c_h = c_{50m}\left(\frac{h}{50}\right)^{\frac{0.37-0.088\log(c_{50m})}{1-0.088\log(50/10)}} \qquad (11)$$

$$k_h = k_{50m}\frac{1-0.088\log(50/10)}{1-0.088\log(h/10)} \qquad (12)$$

Figure 8 shows that the monthly capacity factors oscillate in the 10-35% range and that relatively high values (above 30%) occur between September and January. Based on these results we can estimate the amount of energy (kWh) produced in a month by (13):
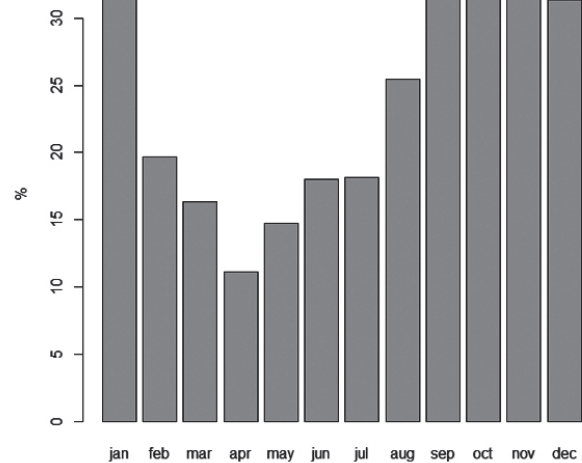
$Energy(month\ i\ \mathrm{MWh}) = Total\ operating\ hours\ in\ month\ i$ x $Rated\ power\ \mathrm{MW}$ x $CF(month\ i)$ (13)

Figure 8 – Capacity factor barplot.


```
windows()
h = 119 # hub heigth (m)
vin=3 # cut in speed (m/s)
vr=12 # rated speed (m/s)
vout=25 # cut out speed (m/s)
cte<-vr^3
cf=numeric(0)
for (i in 1:12) {
   scale=c[i]*(h/50)^((0.37-0.088*log(c[i]))/(1-0.088*log(5)))
   shape=k[i]*(1-0.088*log(5))/(1-0.088*log(h/10))
   aux1<-function(w) (w*w*w*dweibull(w,shape,scale))
   aux2<-function(w) (dweibull(w,shape,scale))
   integral1<-integrate(aux1,vin,vr)
   integral2<-integrate(aux2,vr,vout)
   cf<-c(cf,(1/cte)*integral1$value+integral2$value)
}
names(cf)=c("j","f","m","a","m","j","j","a","s","o","n","d")
barplot(cf*100,ylab="%",cex.axis=2,cex.names=1.5,cex.lab=2)
print(cf)
```

CONCLUSION

In the wind power we can find good examples to teach statistical methods and theory based on real engineering problems to engineering undergraduate students. In this paper we show some examples to a first course on statistics, the examples presented cover the steps of the traditional wind speed data analysis since the data acquisition on internet, statistical analysis up to the procedures to estimate the capacity factor and the amount of monthly energy production. In addition, there are many other topics for advanced courses on Statistics, for example, data cleaning, circular statistics, non-parametric methods, stochastic simulation and time series analysis & forecast.

ACKNOWLEDGMENTS

REFERENCES

AMARANTE, O. A. C. *et al.* **Brazilian wind energy potential atlas**. Brasília, DF: Centro de Pesquisas de Energia Elétrica, 2001.

BLISCHKE, W. R.; SCHEUER, E. M. Tabular aids for fitting Weibull moment estimates. **Naval Research Logistics Quaterly**, v. 33, p. 145-153, 1986.

JANGAMSHETTI, S. H. J.; RAU, V. G. Site matching of wind turbine generators: a case study. **IEEE Transac-**

tions on Energy Conversion, v. 14, n. 4, p. 1.537-1.543, 1999.

JUSTUS, C. G. *et al.* Methods for estimating wind speed frequency distributions. **Journal Applied Meteorology**, n. 17, p. 350-353, 1978.

PALFI, G.; ZAMBON, R. **Hydro and wind power complementarity and scenarization in Brazil**. World Environmental and Water Resources Congress, Cincinnati, 2013.

R CORE TEAM. R: **A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, Available through: <http://www.R-project.org/>. 2014. Accessed 11 November 2014.

REDE SONDA – SONDA - Sistema de Organização Nacional de Dados Ambientais. Available through: <http://sonda.ccst.inpe.br/>. Accessed 31 August 2014.

WWEA - World Wind Energy Association. **Half-year Report 2014**, September, 2014.

VITAE

José Francisco Moreira Pessanha has a BSc in Statistics and BSc/MSc/DSc degrees in electrical engineering. Currently, he is Professor in the Statistics Department at the State University of Rio de Janeiro (UERJ) and researcher at the Brazilian Electric Power Research Center (Eletrobras Cepel) where he has been developing softwares for power systems reliability assessment, load forecasting, regulation of supply quality, design of electricity distribution tariff and wind power forecasting.

Fernando Luiz Cyrino Oliveira is an Assistant Professor of the Industrial Engineering Department of PUC-Rio (Pontifícia Universidade Católica do Rio de Janeiro). He obtained his undergraduate Degree in Industrial Engineering at Federal University of Juiz de Fora and his Master and Doctoral Degrees in Electrical Engineering at PUC-Rio. His research interests include Operations Research, Operations Management, Time Series, Forecasting, Simulation Methods and Literature Reviews. He has published in journals as *European Journal of Operations Research, International Journal of Electrical Power & Energy Systems, Mathematical Problems in Engineering and Journal of Intelligent & Fuzzy Systems*.
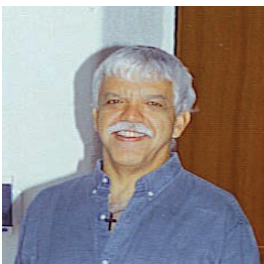
Reinaldo Castro Souza holds PhD in Bayesian forecasting from Warwick University UK. Presently, he is a full Professor in Statistics and Time Series Analysis at Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil and his research interest are on the area of Forecasting with Applications to the EnergyMarket.