The Journal of CPRI,
Vol. 11, No. 3, Sept 2015 pp. 577-584

Regression models in wind power forecasting

Anuradha M*, Keshavan B K**, Ramu T S*** and Sankar V****

Modeling of generation of wind power systems is useful for an effective management and balancing of a power grid, supporting real-time operations. Forecasting the expected wind power production could help to deal with uncertainties. In comparison with the mathematical approach, the data driven approach is useful where both detailed information about the system and real time measurements are unavailable. Winds being a natural phenomenon, statistical methods are more suitable for wind power plants than that of conventional power plants. In this paper, the data on the wind speed and power generated from a location in the state of Karnataka, India, has been analyzed and shown that the probability distribution of wind speed follows Rayleigh or Gaussian/Normal distribution. Short-term wind power forecasting is carried out using Autoregressive models.

Keywords: *Wind power forecasting,rayleigh distribution, gaussian distribution, auto - regression.*

1. 0 INTRODUCTION

The requirement of spinning reserve increases with unforecasted wind power fluctuations. This raises electricity production cost. Thus, it is essential to forecast wind power for effective grid management. The wind power forecast can be effectively carried out by Numerical Weather Prediction (NWP) and statistical models [1].

It is necessary to know the spatial and temporal wind pattern, its velocity and direction reasonably accurately for an efficient location of the wind generators.

At the outset, it is to be understood, that, the wind velocity is a location specific stochastic variable. Of necessity, therefore, one needs to treat it as a random variable conforming to a known distribution [2-3]. This consideration seems, somewhat inappropriate in as much as the random behavior of the wind velocity is not so extremal at that.

Admittedly, the assumption is,the spectrum of velocity data can be seen, by observation, to conform to a Rayleigh distribution model in the time series analysis.

In power system planning, forecasting the possible future wind power generating capacity over a time window subsequent to the observed data, is of particular interest.

This will greatly help load rescheduling besides improving short-term system reliability. The observed wind speed data is location specific and is a discrete random sequence of natural events in time – a time series [4-7]. Forecasting implies a prediction paradigm. Often, a large scatter and

*Professor, PES Institute of Technology, 100ft. Road, BSK III Stage, Bangalore - 560 085, India. E-mail: anuradha@pes.edu, Mobile: +91-9449867802
**Professor and Dean(Academics), PES Institute of Technology, 100ft. Road, BSK III Stage, Bangalore - 560 085, India.
E-mail: keshavanbk@pes.edu, Mobile: +91-9449867801
***Professor,Indian Institute of Science, C V Raman Avenue, Bangalore - 560 012, India. E-mail: proframu8@gmail.com, Mobile: +91- 948043629
****Professor and Director (Foreign affairs and Alumni Matters), Jawaharlal Nehru Technological University Anantapur, Saradha Nagar, Ananthapuramu - 515002, E-mail: vsankar.eee@jntua.ac.in, Mobile: +91-8008802558

hence irreconcilable uncertainties are involved. In the recent past, proven probabilistic methods have been developed to address this issue.

In the recent past, it has been increasingly felt, both from commercial and academic standpoint, to try and develop albeit empirically, models for forecasting of wind power at an epoch of time, beyond the observation window admittedly, this is a difficult task, given the degree of scatter in the location specific wind speeds.

Authors of this contribution have taken the opportunity to try, and apply the existing techniques in time series analysis for forecasting.

2.0 THEORETICAL ASPECTS

In the development of the mathematical model described hereunder, the direction of the wind is seemingly less important, so much so, its scalar counterpart, the wind speed, may conveniently replace the term velocity, a vector quantity. This stems from the fact, that, the wind turbines are designed to be self – seeking / adjusting, in so far as the direction is concerned. It therefore remains to model the location specific wind speed probabilistically.

2.1 Probability Distribution Functions

Let the continuous random variable X, representing the wind speed, be thus v, $X \rightarrow v$, with its domain of definition, $0 \leq X \leq \infty$.

Also, by definition,

$$\Pr[X = v] = f(v) \qquad \text{....(1)}$$

is known to be the probability density function (pdf) of v. The corresponding cumulative distribution function (cdf) $F\ v$ is defined by;

$$\Pr[X \leq v] = F(v)$$

$$F(v) = \int f(v)dv \text{ if it exists,}$$

$$f(v) = F'(v) = \frac{d}{dv}F(v) \qquad \text{....(2)}$$

2.2 The Rayleigh Distribution

A Rayleigh distribution is often found valid when the sub-component of a random variable bears a strong functional relationship with others. One example where the Rayleigh distribution naturally arises is when wind velocity is analyzed into its associative orthogonal vector components; the components themselves are related to their directional cosines. The pdf of a Rayleigh distribution is formally similar to a Gaussian as under;

$$f(x|\sigma) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \qquad \text{....(3)}$$

$$x \geq 0\,,\sigma > 0$$

Where σ is the scale parameter of the distribution.

In fact, the Rayleigh distribution is a degenerate form of a two-parameter Weibull distribution. Rayleigh distribution can also be interpreted as a normal distribution with zero mean and a standard deviation σ, formally represented as N (0, σ).

The cdf of Rayleigh distribution is given by;

$$F(x) = 1 - e^{-\frac{x^2}{2\sigma^2}} \qquad \text{....(4)}$$

$$x \in [0, \infty)$$

2.3 Normal (Gaussian) Distribution

The Normal distribution is among the most widely used probability distribution functions to describe small and often guessable uncertainties. The probability density function (pdf) of a Normal distribution is formally written as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \qquad \text{....(5)}$$

$$-\infty \leq x \leq \infty; \mu, \sigma > 0$$

It is to be noted that the cumulative distribution function (cdf) in this case is not defined in a closed form for all values of the random variable, since $X = x$.

$$F(x) = \int_0^{X=x} f(x)\, dx \quad \text{....(6)}$$

2.4 Parameter Estimation

Several methods of processing random data are as indicated below:

- Least Square Regression (LSR)
- Best Linear Unbiased Estimation (BLUE)
- Best Linear Invariant Estimation (BLIE)
- Monte Carlo Simulation
- Maximum Likelihood method (MLE)

Considering several aspects, the authors pitch upon Auto regression methods in forecasting.

2.5 Probabilistic Modeling of Wind Power

The power contained in wind is given by the kinetic energy of the flowing mass per unit time.

$$P = \frac{1}{2}\rho A v^3 \quad \text{....(7)}$$

Where,

P - Power (W)

ρ - Air density $\approxeq$ 1.225kg/m^3

P - Rotor area (m^2)

v - Wind velocity (m/s) without rotor interference.

Since the stochastic electrical power generated, P, is a power law in v, with an exponent 3 (Cubical law), P needs to be deduced by a transformation process.

2.6 Aspects in wind power Forecasting

The wind power forecasting is seen to be increasingly important in energy system and load management. It also makes a finite contribution to reliability augmentation. The variability in wind power generation is a sequel to irreconcilable uncertainty in the wind speeds, even at a specific location. Forecasting the power output from such stations offer to reduce an unexpected gap between scheduled and actual wind power.

Of the several methods of forecasting currently available, the time series models, Auto Regressive (AR), Moving Average (MA) and combinations thereof have been studied [8-9-10]. In this paper, AR model of relevant order has been seen to adequately handle the stochastic nature of the time series.

2.6.1 Auto regressive model (AR)

Auto regression is essentially a multiple regression mode of the observed data against its own previous values. In this model, the current value of X_t of time series is expressed linearly in terms of previous values X_{t-1}, X_{t-2}, X_{t-3}, and a white noise term in (a_t) with zero mean and variance σ^2.

The *AR(ρ)* model is characterized as

$$X_t = \sum_{i=1}^{p} \emptyset_i Z_{t-i} + a_t \quad \text{....(8)}$$

Where, $\emptyset_i$ = parameters of AR model

With usual notation, *AR(ρ)* refers to the Auto Regressive model of order ρ. It may be noted that orders $\rho=1$ and $\rho=2$ refer to a linear least square and multiple regression respectively. The Eqn. (8) is completely and uniquely defined, if $\emptyset_i$ and a_t can be estimated from the data and its preceding values. The *AR(1)* and *AR(2)* models have been used to estimate the values of $\emptyset_1$, $\emptyset_2$ and σ^2.

2.6.2 Standardized error of estimate

Any time series forecast shall, always be validated against a possible observed value (if it exists). Among the known such procedure is the χ^2 test of significance. However, it is strongly felt that such a test is untenable here. Alternately, a test for adequacy of the forecast in comparison to the observed data has been instituted. One statistic which explains the adequacy is the standardized error of prediction.

The standardized error is computed using the expression;

$$s = \sqrt{\frac{\sum(\hat{X}_0 - \hat{X}_e)^2}{n-1}} \quad ....(9)$$

$$\hat{X}_0 < s < \hat{X}_e \;;$$

Where,

$\hat{X}_0$ - Observed wind power

$\hat{X}_e$ - Estimated wind power. This is of particular relevance since the sample size is relatively small and is biased.

3.0 RESULTS ANALYSIS

The wind velocity data chosen for analysis is drawn from a site in Karnataka, India. Totally 54,000 wind velocity readings are taken for a full year at an interval of 10 min. The wind velocity data was collected for hub heights of 40 m and 50 m and the analysis is carried out on the machine with 40 m hub - height. The data is truncated below cut-in and above furling speed. Analysis of wind speed is carried out by plotting various distribution functions. Wind power forecast for first 7 days of July 2010 and 2011 is done by choosing quarterly data. The probability of wind velocity is plotted to determine the nature of distribution they follow.

3.1 Distribution Functions

The Rayleigh pdf plot (Figure 1) is, as expected skewed towards left and is closer to Gaussian distribution. The Rayleigh cdf (Figure 2) is a special case of Weibull distribution and is often, used in fitting non-extremal data.

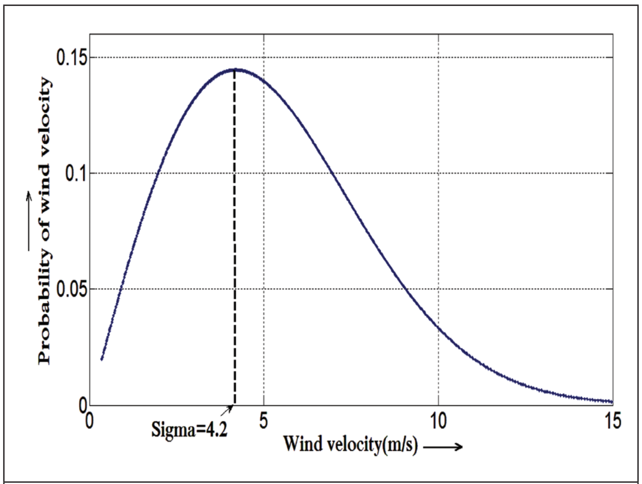

FIG. 1 RAYLEIGH PDF PLOT OF WIND VELOCITY

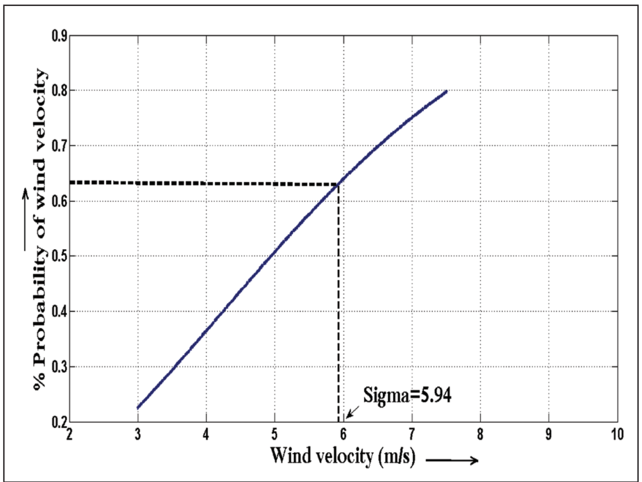

FIG. 2 RAYLEIGH CDF PLOT OF WIND VELOCITY

The scale parameter from the plot is found to be 4.2 m/s which is closer to the average wind velocity of 5.1 m/s as read from the Normal pdf plot (Figure 3) and the standard deviation derived from Normal cdf plot (Figure 4) is 1.92.

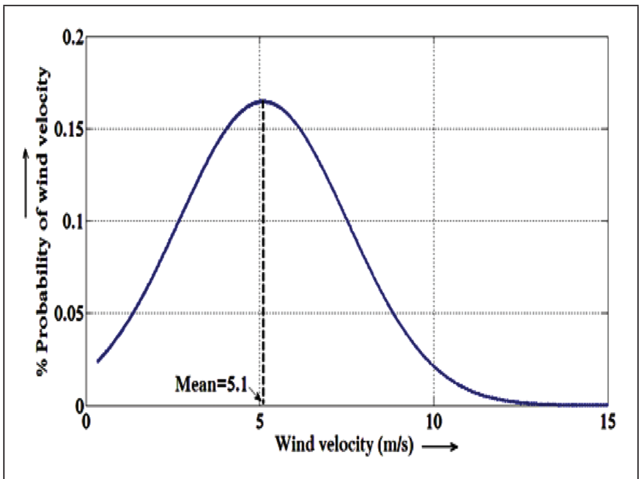

FIG. 3 NORMAL PDF PLOT OF WIND VELOCITY

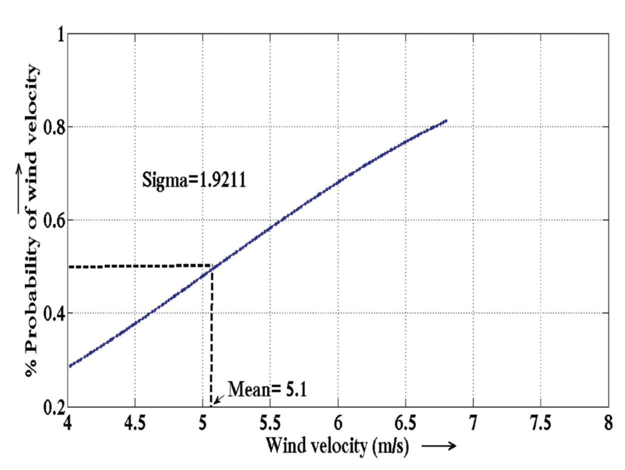

FIG. 4 NORMAL CDF PLOT OF WIND VELOCITY

3.2 Wind Power Forecasting

Long time forecasting of time series is a highly sought after paradigm. However, it is well known that the uncertainties involved in forecasting the time series wind power data beyond the observed time window is only possible if the series in question is stationary in time. Stationarity implies that the variance of the time series remains invariant with respect to the inspection time window, or, the lag, as it is often called. Even when the stationarity guaranteed, the estimation involves large errors in prediction besides associated poor repeatability as well as reproducibility. Authors therefore decided to restrict the predictions to about 7 days beyond the observed period.

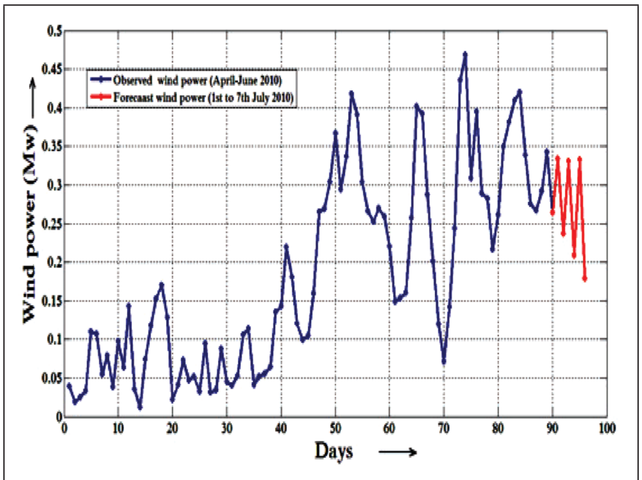

FIG. 5 WIND POWER FORECAST FROM 1ST - 7TH JULY 2010

The observed wind power generated over a period of 90 days of the years 2010 and 2011, sampled daily and the possible power producible over an interval of 7 days beyond the observed period (forecast) has been included in Figure 5 and Figure 6.

The forecast has been made based on the suggested auto regressive model. It can be seen that the forecast seems to be in the expected order. Test for acceptability of the estimates, by way of a standardized error, has been made to verify this aspect. Standardized error is known to be a measure of the variability of estimated data beyond the inspection time window, akin to a biased variance/standard deviation.

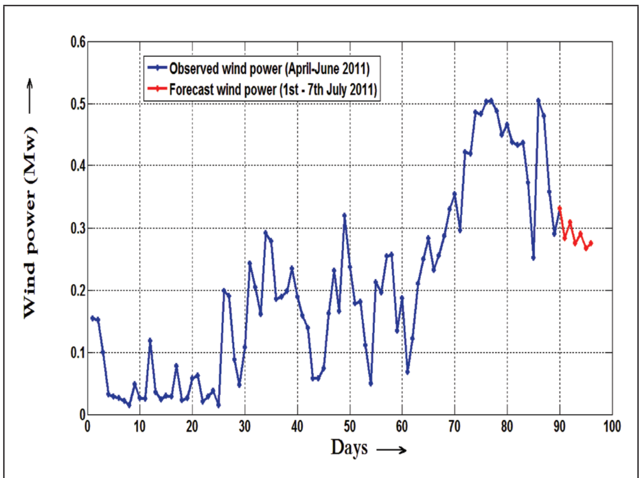

FIG. 6 WIND POWER FORECAST FROM 1ST- 7TH JULY 2011

In the present case, the real time data is available even over the prediction window. So much so, worthwhile comparisons between the two can be made with a reasonable degree of certainty.

As has been said earlier, such a comparison should stand vindicated, albeit statistically, for one to be able to take forward the prediction paradigm for longer intervals.

Figure 7 and Figure 8 compares the forecast value of wind power with the actual data for the period 1-7 days of July, 2010 and 2011. In so doing, the standardized error based on a biased statistics (for small sample sizes) is indicated in Table 1

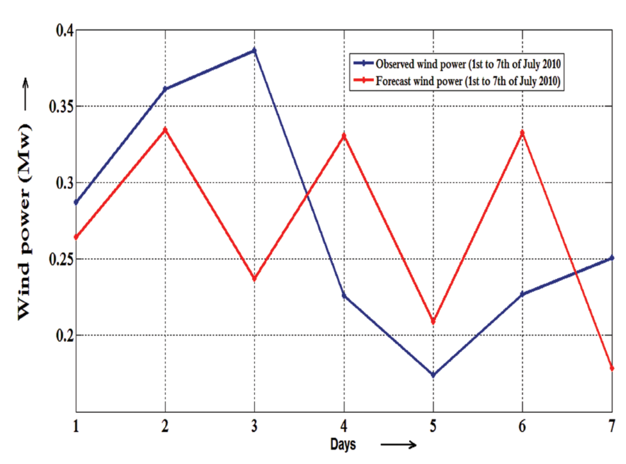

FIG. 7 COMPARISON OF ACTUAL POWER AND FORECAST POWER FROM 1ST - 7TH JULY 2010

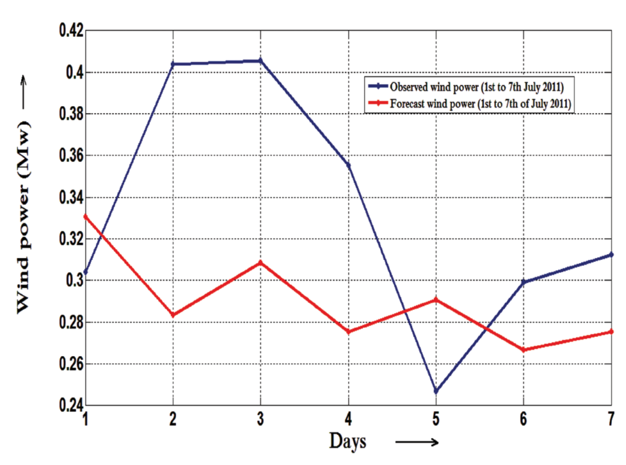

FIG. 8 COMPARISON OF THE ACTUAL POWER AND FORECAST POWER FROM 1ST - 7TH JULY 2011

TABLE 1

STANDARDIZED ERROR ESTIMATION OF WIND POWER GENERATED. (EQN. 9)

Actual Wind Power W (X_0)		**Forecast Wind Power W (X_e)**		**Standard Error of Estimate (s)**	
1st to 7th July 2010	1st to 7th July 2011	1st to 7th July 2010	1st to 7th July 2011	July 2010	July 2011
286980	304139	264317	330610		
361005	405546	334162	283410		
386158	355463	236944	308680		
225737	246645	330571	275280	35196	23157
174005	299188	208692	290610		
226989	312343	332488	266660		
250547	317148	178394	275480		

4.0 DISCUSSION

Literature on this subject is exhaustive. A perusal thereon, suggests that, more often, extremal distributions are invoked in the analysis of wind speed data. However, the degree of variability in the acquired data does not seem to warrant consideration of statistics of extremes.

In discussing the results of forecasting the power time series, two important points have been addressed.

1. Efficacy of the model for forecasting

The second order *AR* model is seen to be reasonably sufficient. However, if there is any evidence of the presence of a Moving Average (*MA*) process of a predictable order, an Auto Regressive Moving Average *ARMA* *(p, q)* is implied. The authors believe that the predicted values are governed by a second order Auto Regressive *AR (2)* process and are seen to be validated by the standardized error. The model is therefore, considered as conforming *AR (2)*

2. The acceptability of predicted values

On the question of the sufficiency of model over the prediction interval, now, apologetically implore the readers that with the suggested model, any further increment in the forecast window is highly uncertain. The wind power forecast is done, by considering the quarterly data from April-June 2010 and 2011.

It may be seen that the standard error computed for the power generation is very much less than the mean of the observed and predicted values and lies within these two ranges. Forecasting of wind-generated power is carried out using a second order Auto Regressive model. The result is validated by computing standardized error. It has been found that the forecast results are applicable for short term forecasting only.

5.0 CONCLUSIONS

The terms of reference of the current research paper was to try and analyze wind speed data in a specified location. In so doing, different distribution functions were examined.

- Among the probability functions considered here, the Rayleigh distribution is seen to be of relevance. As such, this distribution is given a preference.
- The forecasting method described under the relevant section, was a sequel to commercial requirements of power stations. Often, such an information is sought in power system planning and connected reliability studies.
- The *AR(2)* model seems to address this requirement to a reasonable degree of certainty. The statistic, standard error for validation, used by the authors, is seen to be more an adequacy test than the classical χ^2 test of significance.
- The result indicates that the standard error, being much smaller than one mean value, appears to justify the authors' forecasting procedure.
- The authors hasten to add that longer prediction intervals and higher prediction accuracies (with lower uncertainties) can be achieved by modeling the time series using more complex and accommodative expressions. This is an objective for their future work.

REFERENCES

[1] Michael Browser, "Wind energy forecasting, AWS True power", MIT Wind workshop proceedings, Albany, NY, Jan 2011.

[2] R Karki, PoHu and R Billinton, "A simplified Wind Power Generation Model for Reliability Evaluation", IEEE Trans. on Energy Conversion., Vol. 21, No. 2, pp. 533-540, June 2006.

[3] M H Al-Buhairi and AAl- Haydari, "Monthly and Seasonal Investigation of Wind Characteristics and Assessment of Wind Energy Potential in Al-Mokha, Yemen", Scientific Research doi:10.4236/ epe.2012.43017 Energy and Power Engineering, No. 4, pp. 125-131, 2012.

[4] S K Aggarwal, M Gupta, "Wind Power Forecasting: A Review of Statistical Models", International Journal of Energy Science (IJES), Vol. 3, No. 1, pp. 1-11, February 2013.

[5] S Coles, "An Introduction to statistical Modeling of Extreme Values", A volume in the Springer Series in Statistics, Springer-Verlag, London, 2001. ISBN 1- 85233-459-2.

[6] M Milligan, Marc S, Tih-huei Wan, "Statistical Wind Power Forecasting Models:Results for U.S.Wind Farms", WIND POWER 2003, Austin, Texas, pp. 1-17, May 18-21, 2003.

[7] Xiaochen Wang, PengGuo, Xiaobin Huang, "A review of Wind Power Forecasting Models", Energy Procedia 12, pp. 770-778, 2011.

[8] T S Ramu, "Diagnostic Testing and Life Estimation of Power Equipment", Book, New Age Science Limited, 2009.

[9] George E P Box, G M Jenkins and G C Reinsel, "Time Series Analysis Forecasting and Control", Third Edition, Pearson Education. 2003 ed.

[10] J S Armstrong, "Principles of Forecasting: A Handbook for Researchers and Practitioners", (ed.): Norwell, PP18-44, MA Kluwer Academic Publishers, 2001.