Analysis of Curve Fitting for Case Studies for Indian Power Sector

Rajkumar P Narkhede1, Prabha Rastogi1

1Dept. Of Mathematics, J. J. T. University, Churu, Rajasthan, India

Rajkumar.p.narkhede@gmail.com

Abstract. The requirements for Indian electrical power sector is always a challenging task in terms of predication of units of generation of electrical units, load factor (central, state and private) for the power plants, peak demand, peak demand meet and peak demand shortage. The historical data from 1997 to 2018 are used for the analysis. The straight line function and polynomial functions used for the analysis of the data to predict the future. The paper outlines appropriates of the both model in context with the Indian power sector. Further, provide the opportunity to understand the regression coefficient as test condition for the model. The regression coefficients are evaluated for load factor such as 0.9448, 0.9659 and 0.9566 for central, state and private. While regression coefficients for the peak demand and peak demand meet are fairly closer to one. However, the model data values are evaluated and the minimum error in the peak demand is 44% and for peak demand meet is 274% while maximum deviation in peak demand is 107 and for peak demand meet is 743%.

Key Words Regression Coefficient curve fitting straight line method linear regression

1. Introduction
In statistical modelling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables. The most common form of regression analysis is linear regression, in which a researcher finds the line or a curve that most closely fits the data according to a specific mathematical criterion. For example, the method of ordinary least squares compute the unique line that minimizes the sum of squared distances between the true data and that line. To examine certain examples of the above method used by various researchers in various forms to develop models for physical phenomenon. The models can further predict the future values for certain independent variables provided certain dependent variables are known.

A major part of statistical modelling goes into predicting and optimizing current processes to increase profits. In one such study by Zhilin Lu [1] which discusses optimizing wind and photovoltaic energy. The author uses stochastic collocation method for optimizing the wind and photovoltaic energy through IEEE 39 bus as well as 1009 bus systems. The authors conclude that by changing the method of optimization from stochastic economic dispatch (SED) to nested sparse grid-based allocation method (NS-SCM) there is an increase in the economic dispatch of power. This is because the NS-SCM models the process better than SED model. A similar study is carried by Debalina Sahaa et. al. [2], on restructuring the power sector in West Bengal into public sector and private sector utilities is...
estimated. Various parameters are evaluated and their impact on the overall impact on welfare of the people is evaluated. The authors have taken into account various factors like rationalization of tariffs, cross-subsidy, generation, transmission, and distribution losses, etc. By using statistical methods authors have evaluated impact of each change in each parameter to its impact. A study is carried by Jungwoo Leea using levelized cost of electricity (LCOE) for predicting the future investment. Another study is estimating electricity consumption to match generation. This is done to ensure smooth working of the grid system. A study by Yang Zhoua [1] has created a model using 2,66,000 firms in Shanghai to predict their requirements to ensure continuous electricity supply. The algorithm used here is simple linear regression. A comparison of linear regression is done with LASSO, PCA and Ridge regression and it is realized that linear regression is best suited for dimension reduction compared to other algorithms. A study by Anastasia Ioannoua et. al. [4] has found ways in which impact of technological, political, social, environmental factors, etc. is evaluated before the commencement of a renewable energy project. The method assists the stakeholders to take decisions to make power generation sustainable. Semi-quantitative methods like scenario analysis and MCDA are used for addressing non-statistical parameters such as social factors and emergence of competitive technologies. Paraskevas [3] suggests a bottom-up approach for long-term energy planning of the power supply sector integrating mainland and insular electric systems. The model which is developed has a composite cost objective function expressing the NPV of total electricity generation costs over a specific period, comprising annualized investment costs, fixed and variable O&M expenses, fuel costs, GHG emission rights which are all related to power plants plus the cost of electricity imports and the annualized capital costs for the islands-to-mainland interconnections. In conclusion use of such a multi-objective optimization model is not only economical but also better for the environment. Fenghao Cui, et. al.[6] have used cubic smoothing spline, polynomial curve-fitting and nonlinear least squares for fitting models to estimate river water quality. The challenge here is that there are various unknown parameters which are very difficult to model and hence various techniques need to be tried before implementing a specific model. The models were tested by changing the sensitivity index. The variation of model outputs showed a slight difference at a sensitivity index of less than 10% and a significant difference at a sensitivity index of more than 50%. The one-way analysis of variance showed a large p-value of 0.8431, indicating that differences between model data and measured data means are not significant. Thus, the comprehensive application of the numerical curve-fitting techniques of cubic smoothing spline, polynomial curve-fitting, and nonlinear least squares provided a robust estimation procedure to support data analysis and parameter calibration when estimating the unknown parameters of the river water quality management (RWQM). An application of the modelling process is to be able to predict future independent parameters by using historical data. This is done by Farhad Mehmanpazir, et. al. [7] for evaluating supply and demand for future by analysing historical data of past 60 months. A similar research takes place in modelling stock market data to predict which stocks will raise in the future. The critical part here is the choice of various dependent parameters on the values of independent parameters and their weighted influence on the same. Such curve fits work only for a short period of time as the influence of various parameters is not linear. The accurate modelling of the charging characteristics of electric vehicles (EVs) is the basis for the load forecasting, infrastructure planning, and orderly charging management required for proper use of electric vehicles. A study by Zhong Chen et al [9], have done an investigation in Nanjing, China based on operational data available from an EV charging company. On this basis, they carried out stochastic simulation for the load curve of disordered charging and analyzed the potential of the EV charging load participating in the orderly management and its coordination with the output of power generation using renewable energy. Statistical analysis is widely used to predict non-linear phenomenon in physics experiments to derive new laws or to predict certain properties of materials under certain conditions. For example, bending of Aluminium alloys is inelastic in nature and hence it cannot be predicted by natural laws. In a study by Cheng Ming et al [5], Aluminium alloy bending is observed and a model is prepared to estimate the bending in aluminium alloys. Semi-empirical formulas are created which can be used to design Box beams and I beam sections. In evaluating properties of Non-
newtonian fluids (Mohammadhossein Irani \cite{10}), curve for shear stress Vs shear rate is plotted and indexes for temperature and solid volume fraction are evaluated for curve fitting. Thus the model created with two variables, temperature and volume fraction, to evaluate the shear stress/shear strain for other environmental conditions. In GPS systems, there are two important aspects in optimally processing the acoustic measurements: the establishment of the functional model and the refinement of the associated stochastic model. By improving the stochastic model the accuracy of positioning underwater at 3000m depth was increased from 0.426m to 0.306m. (Zhenjie Wang et al \cite{11}). In study, Raman spectra of carbonaceous matter are evaluated automatically by using curve-fitting. This is done as Raman spectra are biased by subjectivity which reduces overall comparability of Raman spectra of carbonaceous material (RSCM). It is concluded that RSCM parameter reflects the metamorphic conditions \cite{12}.

The present paper discusses the appropriateness of the curve fitting tool used for the Indian electrical power generation data with the regression analysis. The history data for the 1997 to 2018 were taken for the analysis and the future is predicated with the different curve fitting functions. The typical errors are evaluated for the various functions used for the data analysis.

2. Results and Discussion

The actual electrical power generation data from year 1997 to 2018 is plotted. Initially, straight line functions for the data analysis. The regression coefficient obtained for the data point is 0.8895 which suggest the better fitting ability. Once the constants obtained for the straight line the model is the used to predict the electrical power generation for each year. It is observed that model and actual electrical power generation is not deviates much. The error in the actual and model electrical power generation capacity is limited within 16.42%. Thus, it is appropriate to use the straight line model to predict the future electrical power generation in Indian context.

![Figure 1. Generation of unit per year from 1997 to 2018](image.png)

\[ y = 32.294x - 64095 \]
\[ R^2 = 0.8855 \]

| Year | Actual unit of Generations | Model Units of Generation |
|------|-----------------------------|---------------------------|
| 1995 | 200                         | 300                       |
| 2000 | 500                         | 600                       |
| 2005 | 800                         | 900                       |
| 2010 | 1100                        | 1200                      |
| 2015 | 1400                        | 1500                      |
| 2020 | 1700                        | 1800                      |

**Table 1.** Load factor for the Indian power plants operated by central, state and private sector.
The load factor data is then plotted with the straight line function for all three power production companies. The constants are evaluated and regression coefficient for central, state and private sector are 0.0205, 0.4925 and 0.1241, respectively (refer figure 2). The model developed and the constants obtained are 0.0205, 0.4925 and 0.1241, respectively (refer Table 1). The model polynomial load factor is no longer the appropriate for the data analysis (refer Table -1).

Table 1 shows the load factor for the power plants operated by the central, state and private power plants. The historical data used for the analysis is from 1997 to 2018. The load factor helps us the aggravator that how much electrical power that could be available for grid planning and distribution. The load factor data is then plotted with the straight line function for all three power production companies. The constants are evaluated and regression coefficient for central, state and private sector obtained are 0.0205, 0.4925 and 0.1241, respectively (refer figure 2). The model developed and the data evaluated for the same it is observed that the variation in the load factor is closer for the central power plant, whereas for state and private power plants model load factor is deviation a large extent within 30 %. Thus, the straight line function may be appropriate to some extent. Same data is used to plot with the forth order polynomial. The regression coefficients are evaluated such as 0.9448, 0.9659 and 0.9566 for central, state and private. The regression coefficients are closer to unity, thereby suggests that the polynomial model works closer with the actual data points. The constants obtained are then used for the model evaluation. It is observed that the model data points deviated in the exuberant manner and the polynomial fit is no longer the appropriate for the data analysis (refer Table -1).
Figure 2. Load factor for the central, private and state power plants from year 1997 to 2018
Figure 3. Energy required and demand meet by the Indian Power sector from 1997 to 2018.

Figure 3 shows the energy availability, demand and shortage for Indian context from year 1997 to 2018. Similar way data analyses are carried out for the all three energy indices with the straight line fit. The constant for the all energy indices are evaluated and the model is fitted for the data. It is observed that the predicted model data points are fairly close to the actual data points. Thereby it is appropriate to use the straight line fit the energy availability, demand and shortage.
Table 2. Peak demand and peak demand meet for the Indian power sector.

| Year | Peak Demand (MW) | Peak Demand Meet (MW) | Model Peak Demand (MW) | Model Peak Demand Meet (MW) | % Model Peak Demand (MW) | % Model Peak Demand Meet (MW) |
|------|-----------------|-----------------------|------------------------|-----------------------------|-------------------------|-----------------------------|
| 1997 | 65343           | 58042                 | 135374                 | 487645                      | 107                     | 740                         |
| 1998 | 67905           | 58445                 | 140449                 | 492897                      | 107                     | 743                         |
| 1999 | 72669           | 63691                 | 145525                 | 498148                      | 100                     | 682                         |
| 2000 | 78037           | 67880                 | 150600                 | 503400                      | 93                      | 642                         |
| 2001 | 78441           | 69189                 | 155675                 | 508652                      | 98                      | 635                         |
| 2002 | 81492           | 71547                 | 160751                 | 513903                      | 97                      | 618                         |
| 2003 | 84574           | 75066                 | 165826                 | 519155                      | 96                      | 592                         |
| 2004 | 87906           | 77652                 | 170901                 | 524407                      | 94                      | 575                         |
| 2005 | 93255           | 81792                 | 175977                 | 529659                      | 89                      | 548                         |
| 2006 | 100715          | 86818                 | 181052                 | 534910                      | 80                      | 516                         |
| 2007 | 108866          | 90793                 | 186127                 | 540162                      | 71                      | 495                         |
| 2008 | 109809          | 96785                 | 191202                 | 545414                      | 74                      | 464                         |
| 2009 | 119166          | 104009                | 196278                 | 550665                      | 65                      | 429                         |
| 2010 | 122287          | 110256                | 201353                 | 555917                      | 65                      | 404                         |
| 2011 | 130006          | 116191                | 206428                 | 561169                      | 59                      | 383                         |
| 2012 | 135453          | 123294                | 211504                 | 566420                      | 56                      | 359                         |
| 2013 | 135918          | 129815                | 216579                 | 571672                      | 59                      | 340                         |
| 2014 | 148166          | 141160                | 221654                 | 576924                      | 50                      | 309                         |
| 2015 | 153366          | 148463                | 226730                 | 582176                      | 48                      | 292                         |
| 2016 | 159542          | 156934                | 231805                 | 587427                      | 45                      | 274                         |
| 2017 | 164066          | 160752                | 236880                 | 592679                      | 44                      | 269                         |

Table 2 shows the peak demand and peak demand meet from 1997 to 2018 for the Indian context. In the earlier discussion it is observed that the straight line function provides better model fit as against the polynomial fit. The straight line fit is used for the analysis of the both power indices. The regression coefficients for the both cases are fairly closer to one (i.e. 0.9845 and 0.9609). The actual data points fairly lie closer to the straight line fit (refer figure 4). However, the model data values are evaluated and the minimum error in the peak demand is 44% and for peak demand meet is 274% while maximum deviation in peak demand is 107 and for peak demand meet is 743% (refer Table 2). Thus, it is observed that always straight line fit for the data under consideration will not meet. The appropriateness for the straight line or polynomial fit is always a matter of the detail investigation and only regression coefficient cannot be a tool provide the information with confidence that the model will work.
3. Conclusions.

The two prominent curve fitting functions used for the electrical power generation indices for the Indian context are analysed. It is observed that the polynomial fit do not yield the model parameters appropriate in the context of the problem under consideration. The straight line functions on the contrary work better for the few data sets whereas for few data sets it deviates to greater extent. Thereby, for straight line function further mathematical better precise conditional parameter is required and only regression coefficient do not provide the appropriateness of the straight line model. In the present case the straight model provides the better and appropriate for fit for the units of electricity generated, while fairly closer to load factor for central own power plants and deviates to greater extent in all other cases.

4. References

[1] Zhilin Lu, Mingbo Liu, Wentian Lu, and Zhuoming Deng 2019 *IEEE-Access* 7 91827
[2] Debalina Saha and Rabindra N. Bhattacharya 2019 *Utilities Policy* 5 62
[3] Yang Zhou, Shuaishuai Zhang, Libo Wu and Yingjie Tian 2019 *Applied Energy* 255 113790
[4] Anastasia Ioannou, Andrew Angusc and Feargal Brennan 2017 *Renewable and Sustainable Energy Reviews* 74 602
[5] Paraskevas N. Georgiou 2016 *Computer and Operation Research* 66 292
[6] Fenghao Cui, Chul Park and Moonil Kim 2019 *Journal of Environmental Management* **249** 109375
[7] Farhad Mehmmanpazir, Kaveh Khalili-Damghani and Ashkan Hafezalkotob 2019 *Resources Policy* **63** 101409
[8] Zhong Chen, Ziqi Zhang, Jiaqing Zhao, Bowen Wu, and Xueliang Huang 2018 *IEEE Access* **6** 24475
[9] Cheng Ming, Shi Yongjiu and Wang Yuanqing 2006 *Tsinghua Science and Technology* **11** 648
[10] Mohammadhossein Irani, Masoud Afrand and Babak Mehmandoust 2019 *Physica A* **531** 120837
[11] Zhenjie Wang, Shuang Zhao, Shengyue Ji, Kaifei He, Zhixi Nie, Huimin Liu and Ru Shan 2019 *Applied Acoustics* **150** 36
[12] N.K. Lünsdorf and J.O. Lünsdorf 2016 *Journal of Coal Geology* **160–161** 51