Stochastic Analysis and Applications
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsaa20

An application of compound probability distributions to electric load modeling

Wiktor Charytoniuk [a] & Joanicjusz Nazarko [a]

[a] Department of Electrical Engineering , Technical University of Bialystok , UL Grunwaldzka 11/15, Bialystok, 15-893, Poland

Published online: 03 Apr 1994.

To cite this article: Wiktor Charytoniuk & Joanicjusz Nazarko (1994) An application of compound probability distributions to electric load modeling, Stochastic Analysis and Applications, 12:1, 31-40, DOI: 10.1080/07362999408809336

To link to this article: http://dx.doi.org/10.1080/07362999408809336

STOCHASTIC ANALYSIS AND APPLICATIONS, 12(1), 31-40 (1994)

AN APPLICATION OF COMPOUND PROBABILITY DISTRIBUTIONS TO ELECTRIC LOAD MODELING

Wiktor Charytoniuk and Joanicjusz Nazarko

Department of Electrical Engineering
Technical University of Bialystok
Ul. Grunwaldzka 11/15
15-893 Bialystok, Poland

ABSTRACT

This article presents a method for expressing uncertainty in network load models using the theory of compound probability distributions, and gives examples for applying the approach to daily and annual peak load models for MV/LV (middle voltage / low voltage) substations.

1. INTRODUCTION

The loads on receiving buses in an electric power network are basic factors determining network design and operation. Network loads are time varying processes of a random nature. Therefore, the load model used for technical and economic network calculations, must express the probabilistic nature of load variation quantitatively. In such load models it is advisable to use stochastic processes or random variables is advisable. The parameters of such models are obtained from data, acquired by observation of load variation and the application of statistical methods.

The deficiency of measured data, to be used in the load models is often very apparent, especially in distribution systems. In order to express the uncertainty in such models statistical methods are applied.

The most renowned methods are probabilistic methods , fuzzy sets theory and the certainty coefficient method.

This article presents a method for expressing the uncertainty in load models by application of the theory of compound probability distributions and gives an example of applying the method to modeling the daily and annual peak loads of receiving MV/LV substations.

2. THE USE OF COMPOUND DISTRIBUTIONS FOR EXPRESSING UNCERTAINTY IN POWER SYSTEM LOAD MODELS

The load of a network bus is most often considered as a time series in discrete form

$$X = \{ X_i \}, \qquad i = 1, \ldots , n.$$

In the general case, a knowledge of n-dimensional probability density function

$$f(x_1, \ldots, x_n | Q),$$

is necessary to the completely describe the probabilistic properties of the series. In the formula, $x_1, \ldots, x_n$ denote realizations of random variables $X_1, \ldots, X_n$, and Q is the matrix of dimension $(r \times n)$ of the distribution parameters.

In many real situations, the time series of a load can be considered as a discrete stationary ergodic process. Sometimes fulfillment of this condition can be achieved by proper choice of the time interval or trough transformations. For example this might be accomplished by consisting the elimination of a periodical component or a trend from a load time series. In this case the load model is reduced to one-dimensional random variable with probability density function $f(x|Q)$ with an r-dimensional parameter vector Q.

In order to account for the uncertainty of model parameters, they are considered as random variables with the specified distribution

$$\xi(q_1, \dots, q_n).$$

In this way, the load model in compound probability distribution form, comprises an uncertainty resulting from both the probabilistic character of load and the inaccuracies in estimating the load model parameters. This total uncertainty can be expressed in the form of a simple distribution, obtained from the compound distribution by the formula

$$f(x) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x|Q)\, \xi(q_1, \dots, q_r)\, dq_1, \dots, dq_r . \quad (1)$$

The distribution of model parameters can be determined by use of a likelihood function, acquired from observations of the load process. The likelihood function is calculated from the equation [2]

$$L(q_1, \dots, q_r) = \prod_{i=1}^{k} f(x_i|Q) \quad (2)$$

where x_i, $i = 1, \dots, k$, imply observed load values.

Data on customers, experts (network operational staff) opinions or prior knowledge from stochastic properties of loads of a given class of buses can be used to determine the preliminary distribution $\xi(q_1, \dots, q_r)$ for the parameters vector Q. In this way the resultant distribution of parameters vector Q is determined, in accordance with Bayes theorem, as the product of a prior distribution and the likelihood function.

3. MODELING PEAK LOADS OF RECEIVING MV/LV SUBSTATIONS FOR UNCERTAINTY

3.1. Probabilistic model of substation peak loads

If $P_d = \{P_{di}\}$, $i = 1, \dots, n$, denotes a sequence of daily peak active loads of substations during the autumn-winter peak time, and i is the succeeding day number for the same period then, in a general a knowledge of the n-dimensional probability density function

$$f(p_{d1}, \ldots, p_{dn}),$$

where $p_{d1}, \ldots, p_{dn}$ imply realizations of random variables $P_{d1}, \ldots, P_{dn}$, is necessary to describe P_d completely.

According to research results, contained in [1,3,4] the process of daily peak loads of a substation during the autumn-winter peak time is a discrete stationary stochastic process with independent values and a normal distribution. Hence, daily peak loads P_{di} can be considered as independent random variables with the same normal distribution, and their probabilistic properties can be expressed directly by a two-parameter probability density function

$$f(p_{d1}|m, \vartheta) = \ldots = f(p_{dn}|m, \vartheta) = f(p_d|m, \vartheta)$$

where m is an expectation and ϑ is a measure of a dispersion of a daily peak load around its average. As a ϑ parameter, a variability coefficient v (a standard deviation divided by an expectation) has been taken, because of its independence of the expectation m. Therefore, the probability density function of a daily peak load has the following form:

$$f(p_d|m, v) = (2\pi)^{-1/2} (mv)^{-1} \exp\{-[(p_d - m)^2 / (2m^2v^2)]\}. \quad (3)$$

In order to take into account the degree of uncertainty related to values of the parameters m i v, they are considered as random variables with the specified distribution $\xi(m, v)$.

An annual peak load (P_a) is also a random variable, which is a function of a daily peak load (P_d), determined as follows:

$$P_a = \max_{i=1, \ldots n} \{P_{di}\} \quad (4)$$

this leads to the relation between distributions of daily and annual peak loads, given by the equation

$$g(p_a|m, v) = n F^{n-1} (p_a|m, v) f(p_a|m, v). \quad (5)$$

where g implies a probability density function of an annual peak load and F a cumulative distribution function of daily peak load.

Equations (3) and (5), under the distribution $\xi(m, v)$ make a probabilistic model of substations daily and annual peak loads.

3.2. Determining a prior distribution of parameters m and v

It is normally assumed [1] that the parameter m has a prior normal distribution with average value m and standard deviation σ_m

$$x(m) = (2\pi)^{-1/2} (\sigma_m)^{-1} \exp\{ - [(m - \mu)^2 / (2\sigma_m{}^2)]\}, \qquad (6)$$

while the parameter v has a gamma distribution with shape coefficient a and scale coefficient b

$$\xi(v) = \frac{b^a}{\Gamma(a)} v^{a-1} \exp(-bv) \qquad (7)$$

where $\Gamma(a)$ is Euler gamma function.

The priori distribution of parameters m and v is determined after obtaining measurements of daily peak loads and after collecting data on customers for a sample from the substation population.

Parameters a and b of the priori distribution of the daily peak load variability coefficient v are calculated on the basis of (acquired from the sample) estimators of the mean E(v) and a standard deviation $\sigma(v)$ of the variability coefficient v, using the following formulas [2]:

$$E(v) = \frac{a}{b}, \qquad \sigma^2 (v) = \frac{a}{b^2} \qquad (8)$$

The prior distribution of daily peak load expectation value m is determined individually for each substation. It is based on accessible customer data supplied from the substations. To this end, on the basis of sample from a substation population, a regression model, expressing the correlation between parameter m and a set of customers features (explanatory variables), existing in the substation population, is determined. The resulting model is used to evaluate the average value μ and the standard deviations σ_m of the parameter m for the given substation as function of the vector of explanatory variables Z'. The value of parameter μ is calculated by the equation

$$\mu = E(m) = Z' A \tag{9}$$

as an expectation m forecast, while the value of $\sigma_m{}^2$ is given by the equation

$$\sigma_m{}^2 = s^2 \left[1 + Z'^T \cdot (Z^T Z)^{-1} Z'\right] \tag{10}$$

as a variation of the forecast. In equations (9) and (10) A implies a vector of multiple regression coefficients, obtained using the method of least squares [2], Z is a matrix of explanatory variable observations and s^2 is an estimator of residual variations.

In cases where customer data is lacking, it is assumed that the parameter m has a prior uniform distribution.

Considering a possibility of the existence of a correlation between some of the explanatory variables, it is necessary to use a method of elimination of these variables, which can be excluded from the model without an essential deterioration of the model quality, for instance the rejection method [2].

Parameters m and v are statistically independent, so that their joint (two-dimensional) prior distribution has a form given by the equation

$$\xi(m, v) \ \alpha \ \exp\left\{-\frac{(m-\mu)^2}{2\sigma_m{}^2}\right\} v^{a-1} \exp(-bv)\,. \tag{11}$$

3.3. Determining a posterior distribution of parameters m and v

Subsequent measurements of substation daily peak loads are used to update the distribution of parameters m and v. The posterior distribution of parameters m and v, after regard to the daily peak loads sample of size k, is calculated by the formula:

$$\xi(m, v|p_{d1}, \ldots, p_{dk}) \ \alpha \ \xi(m, v)\, L(m, v|p_{d1}, \ldots, p_{dk}). \tag{12}$$

where $L(m, v|p_{d1}, ..., p_{dk})$ is the likelihood function of parameters m and v, associated with the sample $p_{d1}, ..., p_{dk}$.

The likelihood function for variable P_d of the given distribution $f(p_d|m, v)$ has the form:

$$L(m, v|p_{d1}, ..., p_{dk}) \propto (mv)^{-k} \exp\left[-\frac{k(m-\bar{p}_d)^2 + (k-1)\hat{\sigma}^2(p_d)}{2m^2v^2}\right] \tag{13}$$

where $\bar{p}_d$ and $\hat{\sigma}^2(p_d)$ imply estimators of the average value and the variance of daily peak load, determined on the basis of the $p_{d1}, ..., p_{dk}$ sample of size k.

Substituting the equation (13) into (12), the following form of the density function of the posterior distribution of parameters m and v is obtained:

$$\xi(m, v|p_{d1}, ..., p_{dk}) \propto \mu^{-k} v^{a-1-k} \times$$

$$\times \exp - \left[bv + \frac{(m-\mu)^2}{2\sigma_m^2} + \frac{k(m-p_d)^2 + (k-1)\sigma^2(p_d)}{2m^2v^2}\right]. \tag{14}$$

The constant, associated with posterior density function of parameters m and v, is calculated from the condition

$$f(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \xi(m, v|p_{d1}, ..., p_{dk})\, dm\, dv = 1. \tag{15}$$

4. A COMPUTATIONAL EXAMPLE

An application of the proposed method of modeling peak loads of MV/LV substations is presented by an example of substations where the sample size is 16. These substations supply rural customers from Bialystok Power Utility region [3].

Data for loads and customer features for the substations are used to build a peak load model. On the basis of this data the following are

determined using the method described in section 3:
1) The prior distribution of daily peak load variability coefficient v for the population;
2) Equations which enable the calculate of the average values μ and the variance σ_m^2 of the prior forecast of a daily peak load expectation value, for each substation from the population, on the basis of the available customer data.

Parameters for the distribution of the variability coefficient in the population, as calculated from formulae (6), amount to:

$$a = 17.8, \qquad b=85.0 .$$

Equations for calculating parameters μ and σ_m^2 are determined as follows:
1) From the customer data set, we obtain the parameters for the regression model, which expresses the correlation between parameter m and the customer features sets. These are estimated, using the rejection method [2]. The set of customer data comprises an annual energy consumption - A (MWh), an installed power - P (kW), a number of customers - n, and an area of arable land - S (ha);
2) The results of the p.1 model are confirmed by calculating the multiple correlation coefficient and checking the multiple regression significance using F-Snedecer's test [2];
3) From the set of significant data and equations (9) and (10) we can evaluate parameters μ and σ_m^2. This gives the priori distribution for each substation, on the basis of customer data.

The results of above calculations are presented in Table 1.

The priori model of the load of a particular substation based upon calculated parameters can be updated according to the equation (14), using available measurements of daily peak loads.

5. CONCLUSIONS

Application of compound probability distribution theory to the modeling of power network bus loads enables us to express the uncertainty of both the

TABLE 1
Formation of equations for calculating parameters of a prior distribution of an expectation value of substation daily peak loads

Explanatory variables		P_i, A_a, n_c, n_i, S
Significant explanatory variables		P_i, A_a
Multiple correlation coefficient		0.931
Test of significance of multiple regression	F	35.9
	F_{crit}	4.05
	result	+
Standard deviation of residuals		2.752
Equation for calculating parameter μ [kW]		$6.26 + 0.0219 P_i + 0.111 A_a$
Equation for calculating parameter σ_m^2 [kW^2]		$12.56 - 0.0148 P_i - 0.154 A_a - 371\times10^{-6} P_i A_a + 371\times10^{-6} P_i^2 + 944\times10^{-6} A_a^2$

probabilistic character of the loads and the impossibility of accurate estimation of load model parameters.

Using Bayes theorem to determine the distribution of the load model parameters allows us to effectively and flexibly use our prior knowledge of stochastic properties of the network bus loads, customers data, and load measurements to formulate accurate load models.

ACKNOWLEDGMENTS

Research for this article was supported in part by a grant from the International Research & Exchanges Board (IREX) with funds provided by the Philip D. Reed Foundation, Inc. and the United States Information Agency. None of these organizations is responsible for the views expressed.

REFERENCES

[1] W. Charytoniuk: Estimation of Peak Loads of Rural MV/LV Transformer

Substations Using Bayesian Statistical Inference Method. Ph.D. Thesis. Technical University of Warsaw, 1991.

[2] N.L. Johnson, F.C. Leone: Statistic and Experimental Design in Engineering and the Physical Science. John Willey and Sons Ltd., New York
1977.

[3] J. Nazarko: State Estimation of Power Distribution Networks. Technical University of Bialystok, Bialystok 1991.

[4] H.L. Willis, C.L. Brooks: An Interactive End-Use Electric Load Model for Microcomputer Implementation. IEEE Trans. on Power Apparatus and Systems. Vol. PAS-102, 1983, No. 11.