

# Transmission Loss Modelling and Analysis with Multiple Linear Regression

S. Appalasamy<sup>a,b</sup>, H.S. Gan<sup>a</sup>, O.D. Jones<sup>a</sup>, N.H. Moin<sup>b</sup>, C.S. Tan<sup>c</sup>

<sup>a</sup>Department of Mathematics and Statistics, University of Melbourne, 3010, Melbourne, Victoria, Australia

<sup>b</sup>Institute of Mathematical Sciences, University of Malaya, 50603, Kuala Lumpur, Malaysia

<sup>c</sup>Institute of Energy Policy and Research (IEPRE), University Tenaga Nasional (UNITEN), 43009, Jalan Kajang-Puchong, Kajang, Selangor, Malaysia  
Email: [srekham3@yahoo.com.my](mailto:srekham3@yahoo.com.my)

**Abstract:** Unit commitment (UC) and economic dispatch (ED) are two crucial optimisation problems in the short term operational planning of power systems. For a given scheduling period, UC determines the optimal set of generating units to be in service whereas ED determines the economic distribution of generation values for a known set of generators. Both of these problems are modelled as aggregated supply and demand problems, and require an estimate of the transmission loss. Therefore the accuracy of the approximated transmission loss within these problems is vital in ensuring the optimality and feasibility of the solutions. The increasing penetration of renewable energy (RE) technologies into the grid has increased the volatility of the transmitted power, making it harder to approximate the transmission loss using existing techniques. A robust and reliable approximation is required, valid across a wide range of transmission values.

Consider a power network with a set of nodes connected by transmission lines, with subset  $B$  of nodes with demand and subset  $N$  of nodes with generators. Let  $d_i$  be the real power demand at node  $i \in B$ ,  $p_j$ , the real power generated at node  $j \in N$  and  $L$ , the total real power transmission losses in the system. Without loss of generality let generator node 0 be the slack bus and write  $N_0 = N \setminus \{0\}$  for the generation nodes excluding the slack bus. This paper looks into a new way of modelling the aggregated transmission loss, using multiple linear regression. The fitted model's form is

$$L(k) = \sum_{i \in N_0} \sum_{j \in N_0} \alpha_{ij} p_i(k) p_j(k) + \sum_{i \in B} \sum_{j \in B} \gamma_{ij} d_i(k) d_j(k) + \sum_{i \in B} \sum_{j \in N_0} \eta_{ij} d_i(k) p_j(k) + \epsilon(k)$$

where  $k = (1, \dots, n)$  is the observation number,  $\epsilon(k)$  is the error and  $\alpha_{ij}$ ,  $\beta_{ij}$  and  $\eta_{ij}$  are coefficients fitted using least squares. The proposed model does not rely only on a particular base case and does not make simplifying assumptions, as seen in previous models, though we do assume that the topology of the power network does not change. This makes the model more robust than existing approximations.

In this paper the effect of power demand (load) at each demand point, power generation and voltage magnitudes for each generator are tested for eight different scenarios created using J.H. Chows 3-Machine 9-Bus benchmark problem which is quoted in Zimmerman et al. (2011). In each scenario we compare our proposed model with loss approximation models currently used in industry. From the analysis we see that our proposed model outperforms the existing models, and gives good approximations for a wide range of inputs. We also show that the performance measures used to compare the models can be used to determine a best base case.

Finally, we show that by looking at the effect of voltage on how well our model fits, we are able to determine voltage limits for generators that are best, in the sense that they minimise the instability caused to load flows due to improper voltage magnitude values.

**Keywords:** Transmission loss modelling, Multiple Linear Regression, Performance Measures, Voltage Instability

## 1 INTRODUCTION

High penetration of renewable generation (i.e. wind and solar) posts challenges to power system operators in grid management and generation scheduling. The inherent intermittency and variability of renewable resources such as wind and solar require that current industry practices, such as unit commitment (UC) and economic dispatch (ED), be advanced to accommodate large amounts of renewable generation (Wang et al., 2009). UC is the problem of optimal scheduling of generating units to be in service for a given period of time, and also for how long they are to be in service, whereas ED is the problem of determining an economic distribution of the generation for a given schedule in a static environment. One of the major uncertainties in these optimisation problems are the transmission losses. The transmission losses satisfy a system of non-linear equations, the solving of which takes too long to be incorporated into the solution of real time UC or ED problems. Thus good approximation of the transmission losses is crucial, in particular estimation of the total power loss, especially for a network with long distance power transmission lines or cases with vast area but low load density.

Two approaches to approximating the loss are commonly used (Kothari and Dhillon, 2004; Jiang, 1995; Nanda et al., 1999; Haiwang et al., 2013). The first uses an approximation based on a particular base case, which can be solved exactly (beforehand). Provided the state of the system, in particular the current aggregated power demand ratio, is close to the base case this works well. Unfortunately when highly variable inputs from renewable energy sources are included, the state of the system can quickly move away from the current base case, rapidly degrading the quality of the approximation. We will use this approach to compare to our proposed model. The second is to use generalized generation shift distribution factors that are formulated through linear approximations to the power flow equations, which can be solved quickly. This approach has the advantage that it gives approximate transmission loss for each line, unlike the first approach which only gives total transmission loss, but unfortunately it has been shown to give a poor approximation in general (Jiang, 1995).

In this paper, we show that a robust and reliable approximated transmission loss function can be developed using multiple linear regression on a set of simulated or historical load flow data. This method does not depend on a particular base case nor does it make any simplifying assumptions about transmission losses, though we do assume that the topology of the power network is kept constant. The model is an empirical model, meaning that it is not derived theoretically but rather fitted to a set of observed data. The data can either be from historical observations of the network, or from exact simulations of the network. Exact network simulations are time consuming, but only need to be carried out once: once we have fitted our regression model we can use it without needing further exact simulations.

The effectiveness of our proposed method is tested using a set of scenarios based on the 3-Machine 9-Bus benchmark problem by Chow quoted in Zimmerman et al. (2011), and we compare it with the approach of Dopazo (Kothari and Dhillon, 2004), which uses an approximation based on an exact solution for a given base case. The quality of our approximation can be quantified using statistical performance measures such as the coefficient of determination and the mean squared error. Our proposed method always performs better than the approach of Dopazo (Kothari and Dhillon, 2004). Moreover, we show that our statistical performance measures can be used to determine the best base case to use in the approach of Dopazo (Kothari and Dhillon, 2004), in which case it performs nearly as well as our method.

The paper develops with Section 2 describing the exact transmission loss equations, and the approximation approach of Dopazo (Kothari and Dhillon, 2004). Section 3 gives the proposed model, then Section 4 applies our model to eight different scenarios based on our benchmark problem, and compares it to the approach of Dopazo and also its modified version which represents the loss approximation function used generally in industry (Kothari and Dhillon, 2004; Murty, 2009). Finally in Section 5 we give some conclusions.

## 2 POWER FLOW AND LOSSES

Suppose that our power network consists of a set of nodes (busses)  $\Omega$ , connected by transmission lines. Existence of resistance and reactance on each transmission line causes losses to be accrued as power flows. We suppose that on a certain subset of nodes  $B$  there is a known demand, and there is a subset of nodes  $N$  where we can generate power. Let  $d_i$  be the real power demand at node  $i \in B$  and  $p_j$  be the real power generated at node  $j \in N$ . Let  $L$  be the total real power lost in transmission, then we must have  $L = \sum_{j \in N} p_j - \sum_{i \in B} d_i$ . Because  $L$  depends on the  $p_j$  and  $d_i$ , this equation tell us that if we know the demands and the power generated at all but one generators, then in a feasible solution the power at the remaining generator is determined. Thus it is normal to identify one of the generator nodes as a *slack bus*, and suppose that the power generated at the

slack bus is a function of the demands and power generated elsewhere. We will denote the slack bus by 0, and let  $N_0 = N \setminus \{0\}$  be the generation nodes excluding the slack bus.

The exact loss on the transmission line from  $i$  to  $j$ , for  $i, j \in \Omega$ , is

$$L_{ij} = S_{ij} + S_{ji}$$

where  $S_{ij}$  is the apparent power in the  $ij$  direction given as

$$S_{ij} = V_i((V_i^* - V_j^*)Y_{ij}^* + V_i^*Y_{ij0}^*)$$

with  $V_i$  being the complex voltage value at a bus  $i$ ,  $V_i^*$  its conjugate,  $Y_{ij}$  the admittance of line  $ij$  and  $Y_{ij0}$ , the shunt admittance.

Kothari and Dhillon (2004) showed that the total loss can be written as

$$L = \sum_{i \in \Omega} \sum_{j \in \Omega} L_{ij} = \sum_{i \in B} \sum_{j \in B} d_i B_{ij} d_j - \sum_{i \in B} \sum_{j \in N} (B_{ij} + B_{ji}) d_i p_j + \sum_{i \in N} \sum_{j \in N} p_i B_{ij} p_j \quad (1)$$

where

$$B_{ij} = \frac{\cos(\theta_i - \theta_j)}{|V_i||V_j| \cos(\phi_i) \cos(\phi_j)} R_{ij}.$$

Here  $R_{ij}$  is the real impedance on line  $ij$ ,  $\phi_i = \arctan(P_i/Q_i)$ ,  $P_i$  is the real power at node  $i$  (so this would be  $p_i$  if  $i \in N$  or  $-d_i$  if  $i \in B$ ),  $Q_i$  is the reactive power at node  $i$ , and  $\theta_i$  is the electricity current phase angle at node  $i$ . (See also Murty, 2009.)

It is seen that coefficients  $B_{ij}$  can only be calculated once all  $p_j$  are known, which means that  $L$  has to be known too. None-the-less we can use Equation (1) to obtain a quadratic approximation to  $L$ , by calculating the  $B_{ij}$  for a fixed base case, and then using these fixed  $B_{ij}$  as the  $p_j$  vary. We call this a Dopazo type model. Clearly for this to work, we require that the most suitable base case problem is chosen for the future predicted system, as a wrong choice of base case (i.e. a base case that differs significantly from the actual system state), would give a bad prediction. This approximation is generally only reasonable if the voltages, phase angles, resistance/reactance ratios, and relative loads all remain fixed (so total load can vary within these constraints). In our experience Dopazo type approximations are particularly sensitive to changes in the relative loads at each node, which makes their application especially challenging for a power system interconnected with RE, as its intermittent generation causes a volatile environment.

### 3 MULTIPLE LINEAR REGRESSION (MLR) FOR AGGREGATED TRANSMISSION LOSS

Our model is an empirical model for  $L$ , taking as inputs the  $d_i$  and  $p_j$ , but not depending explicitly on the  $V_i$ ,  $\theta_i$  or  $\phi_i$ . Using a particular test case, we will show that we can approximate  $L$  with good accuracy for a wide range of  $d_i$ ,  $p_j$ ,  $V_i$ , and  $\phi_i$ , spanning the limits of the whole system. Once fitted, the model does not need to be adjusted unless the network topology changes.

To fit the model we need data. We will suppose that for  $k = (1, \dots, n)$  we have observations of the demand, generation and loss, that is we observe

$$d_i(k) \text{ for } i \in B, \quad p_j(k) \text{ for } j \in N_0, \quad \text{and } L(k).$$

It is not assumed that  $V_i(k)$ ,  $\theta_i(k)$  and  $\phi_i(k)$  are constant in  $k$ , so it is possible to have different values of  $L$  for the same values of  $d_i$  and  $p_j$ . The data can come from past observations of the system we are modelling, or from a numerical simulator such as MATPOWER 4.1, (Zimmerman et al., 2011), which is what we used for the test case described in the next section. In either case it is necessary that the data covers the entire range of states that the system will operate in.

Our model has the form

$$L(k) = \sum_{i \in N_0} \sum_{j \in N_0} \alpha_{ij} p_i(k) p_j(k) + \sum_{i \in B} \sum_{j \in B} \gamma_{ij} d_i(k) d_j(k) + \sum_{i \in B} \sum_{j \in N_0} \eta_{ij} d_i(k) p_j(k) + \epsilon(k) \quad (2)$$

where  $\epsilon(k)$  is the error we get using this quadratic form. This is an example of a multiple linear regression (MLR), and the coefficients  $\alpha_{ij}$ ,  $\beta_{ij}$  and  $\eta_{ij}$  can be fitted using least squares. Once fitted we do not need to

recompute the coefficients unless there is a change in the whole network's topology. This model's structure is similar to (1), but note that to ensure that the model is non-singular we have removed the dependent  $p_0$  terms, corresponding to the power generated at the slack bus.  $p_0$  can always be recovered since

$$p_0 = \sum_{i \in B} d_i - \sum_{i \in N_0} p_i + L.$$

How well the model fits can be quantified using for example the coefficient of determination and mean squared error, and we can also use these to compare the model to alternatives.

#### 4 NUMERICAL COMPUTATION AND ANALYSIS

To test our model we used a standard benchmark network 3-Machine 9-Bus system (Zimmerman et al., 2011), consisting of nine nodes, three of which have generators attached. The first generator node was chosen as the slack bus. Using this network we developed eight scenarios, depending on how we restricted the demands and voltages. For each scenario we generated a data set of 1000 observations using the MATLAB package MATPOWER 4.1 (Zimmerman et al., 2011). This is a simulation tool for solving power flow and optimal power flow problems. Scenario 1 avoids any assumptions about loads, voltage magnitudes or power factors,  $\cos(\phi_i)$ , allowing all of these to vary. Scenarios 2, 3 and 4 restrict the voltage to its mid, min and max values respectively. These scenarios provide us an insight into the effect of the voltage on the stability of the power system in a volatile load environment. Scenarios 5–8 impose the assumption made by many past models, namely that the ratio of the individual loads over the total load stays fixed. This allows us to compare our model to Dopazo type models in a situation favourable to the latter. The power factors for all scenarios are varied strictly between 0 and 1.

For each scenario a set of 1000 observations of  $p_i$ ,  $d_i$  and  $|V_i|$  was used. We then fitted the model (2) and calculated two goodness of fit measures, the coefficient of determination  $R^2$ , and the mean squared error MSE. Let  $\hat{L}(k)$  be our fitted approximation to  $L(k)$ , then

$$R^2 = 1 - \frac{\sum_k (L(k) - \hat{L}(k))^2}{\sum_k (L(k) - \bar{L})^2} \in [0, 1]$$

$$\text{MSE} = \frac{1}{n} \sum_k (L(k) - \hat{L}(k))^2$$

where  $\bar{L}$  is the average loss. We denote this model MLR.

We next compared our model to an optimised Dopazo model. Here we constructed 1000 models, choosing each data point in turn as the base case. Given a base case, we used it to calculate the coefficients  $B_{ij}$ , then keeping the  $B_{ij}$  fixed the loss was approximated using (1), and the  $R^2$  and MSE calculated as above. We then chose the base case that minimised these. That is, we chose the base case that gave the best approximation as measured by high  $R^2 \approx 1$  and low  $\text{MSE} \approx 0$ . It turns out that in our example the base case that produced the maximum  $R^2$  also gave the minimum MSE each time. We denote this model DM.

We also compared our model to a restriction of the Dopazo model corresponding to the approach commonly used in industry (Kothari and Dhillon, 2004; Murty, 2009). Here we again construct 1000 models, choosing each data point as a base case and calculating the  $B_{ij}$  for that case. However, when we use (1) to estimate the loss, we don't allow the demand  $d_i$  to vary, but keep these values equal to their values at the base case. That is, we are using just the generated power  $p_j$  as inputs into the approximation, and not the demands  $d_i$ . Again we choose the best approximation according to the  $R^2$  and MSE, and we denote this model D.

Our test scenarios were designed to test how robust the MLR and Dopazo models were to variation in the voltage magnitudes, demand, demand aggregated ratio and power factor. All our scenarios used the same power network topology and therefore the real values of the impedance matrix,  $R_{ij}$ , were constant, while  $d_i$ ,  $p_j$ ,  $V_i$ ,  $\theta_i$  and  $\phi_i$  were varied. Table 1 reports the results of the performance measures  $R^2$  and MSE for all 3 models in all 8 scenarios. We see that the  $R^2$  value and the MSE value of the MLR are the best for all given scenarios.

It is to be noted that the D model performs best when the ratio of the aggregated demand over total demand is kept constant. Its  $R^2$  value is always larger than 0.9 and its MSE value is between 0.7 and 0.3. These values improve when the voltage magnitude is kept constant and are the best when  $|V_i| = 0.9$ . However for a volatile

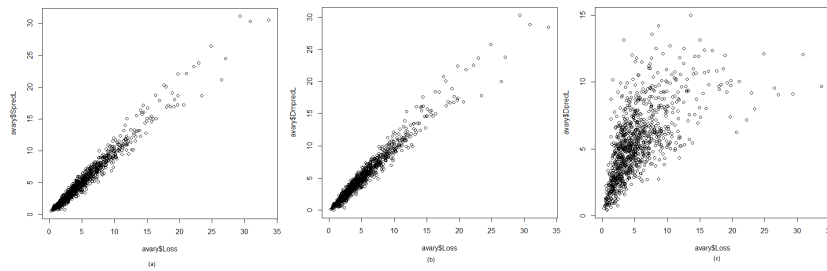
demand ratio its performance reduces drastically. It is also to be noted that in a volatile environment the best MSE value is observed when  $|V_i| = 1.1$ .

The MLR method gave  $R^2$  values all close to 1, though its fit was poorest when the  $|V_i|$  varied. Its also noted that for both constant and non constant  $d_i/\sum_{i \in B} d_i$ , the best performance was seen when  $|V_i| = 1.1$ . Comparing the results of the DM model to the MLR model, it is seen that the DM model’s performance is approximately the same as the MLR’s, though the MLR still performs a little better for all cases. It is also to be noted that the DM model performs best when  $|V_i| = 1.1$ .

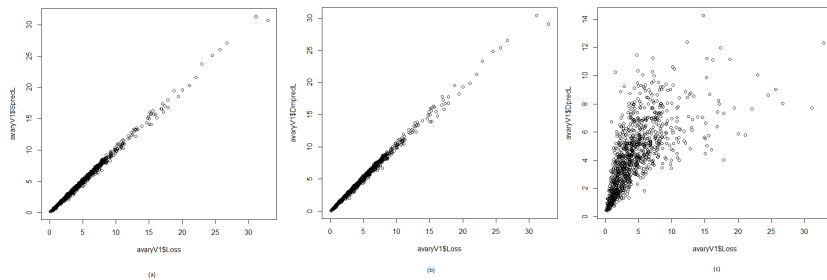
How well the three models fit can also be judged by plotting predicted losses against exact losses, as in Figures 1–6. Larger values of  $R^2$  and smaller values of mean squared error correspond to a stronger linear relationship with a slope of 1. From these figures we also see that the given benchmark problem is more stable, that is the loss is more regular, when voltage magnitudes are kept close to the value of 1.1 (scenarios 4 and 8).

**Table 1.** Results (V indicates varying and C constant)

| Scenario | $d_i/\sum d_i$ | $ V_i $ | $R^2_{MLR}$ | $MSE_{MLR}$ | $R^2_D$ | $MSE_D$ | $R^2_{DM}$ | $MSE_{DM}$ |
|----------|----------------|---------|-------------|-------------|---------|---------|------------|------------|
| 1        | V              | V       | 0.965       | 0.609       | 0.408   | 10.147  | 0.958      | 0.713      |
| 2        | V              | 1.0     | 0.996       | 0.062       | 0.425   | 8.735   | 0.994      | 0.094      |
| 3        | V              | 0.9     | 0.988       | 0.284       | 0.405   | 14.268  | 0.981      | 0.448      |
| 4        | V              | 1.1     | 0.998       | 0.02        | 0.436   | 5.433   | 0.997      | 0.028      |
| 5        | C              | V       | 0.952       | 0.459       | 0.931   | 0.675   | 0.946      | 0.527      |
| 6        | C              | 1.0     | 0.995       | 0.076       | 0.921   | 0.703   | 0.993      | 0.115      |
| 7        | C              | 0.9     | 0.995       | 0.045       | 0.988   | 0.0471  | 0.994      | 0.054      |
| 8        | C              | 1.1     | 0.997       | 0.010       | 0.968   | 0.282   | 0.996      | 0.013      |



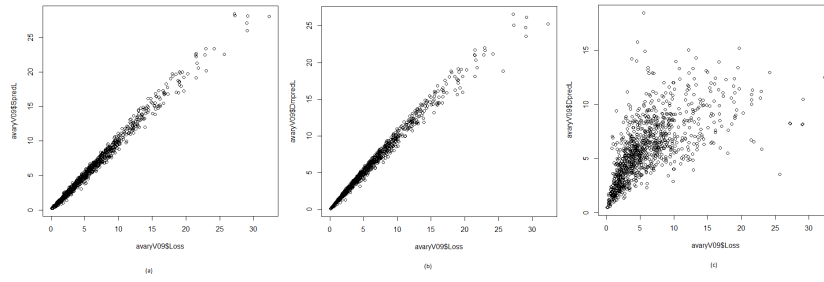
**Figure 1.** Graphs illustrating the prediction accuracy of (a) MLR, (b) DM and (c) D loss model for scenario 1



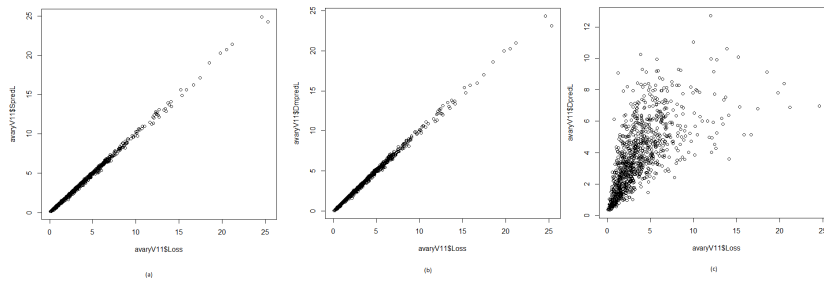
**Figure 2.** Graphs illustrating the prediction accuracy of (a) MLR, (b) DM and (c) D loss model for scenario 2

### 5 CONCLUSIONS

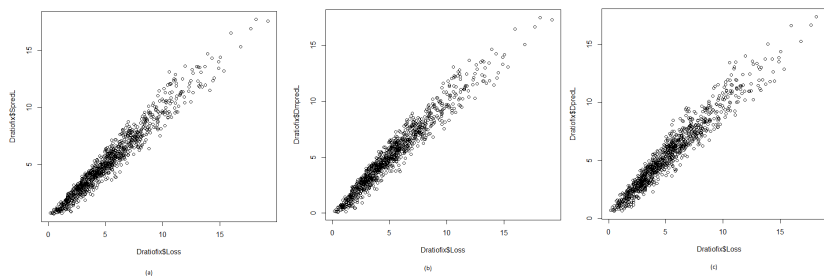
This paper proposed a new model for approximating the transmission loss using multiple linear regression. It was seen that model outperformed the Dopazo model on a given test network. The robustness of the model was



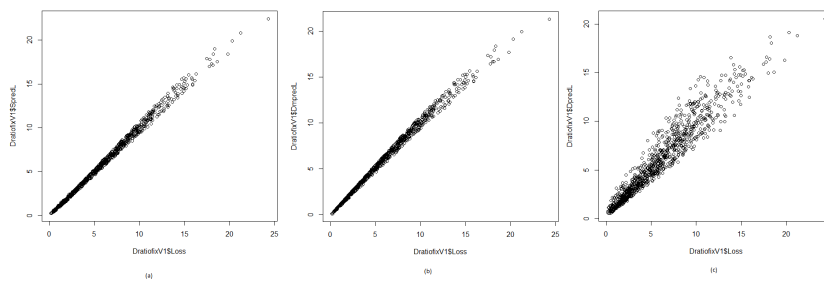
**Figure 3.** Graphs illustrating the prediction accuracy of (a) MLR, (b) DM and (c) D loss model for scenario 3



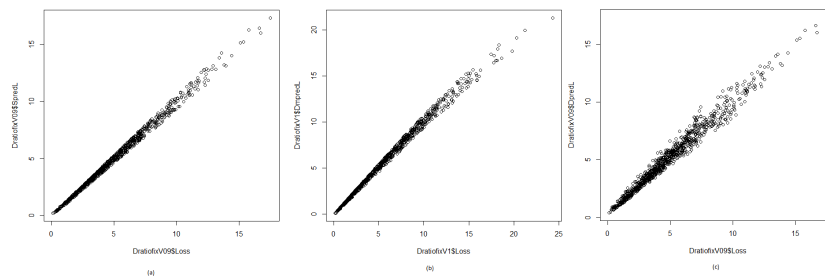
**Figure 4.** Graphs illustrating the prediction accuracy of (a) MLR, (b) DM and (c) D loss model for scenario 4



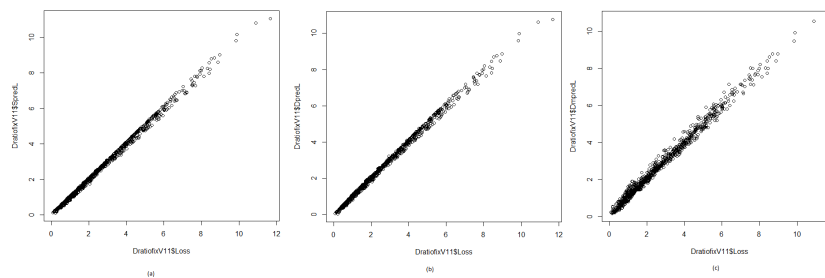
**Figure 5.** Graphs illustrating the prediction accuracy of (a) MLR, (b) DM and (c) D loss model for scenario 5



**Figure 6.** Graphs illustrating the prediction accuracy of (a) MLR, (b) DM and (c) D loss model for scenario 6



**Figure 7.** Graphs illustrating the prediction accuracy of (a) MLR, (b) DM and (c) D loss model for scenario 7



**Figure 8.** Graphs illustrating the prediction accuracy of (a) MLR, (b) DM and (c) D loss model for scenario 8

reflected through good performance measures and scatter plots for a variety of scenarios with volatile demand. It was also shown that the mean squared error and coefficient of determination could be used to determine a good base case problem for the Dopazo model, however we actually get a better fit with less effort by fitting our quadratic model using least squares. Finally, we showed that the MLR and Dopazo models could also be used to suggest suitable voltage limits for a given power system.

#### ACKNOWLEDGEMENT

The first author would like to thank the Malaysian Ministry of Higher Education for funding her doctoral studies at the University of Melbourne.

#### REFERENCES

Haiwang, Z. Q. X., Y. Wang, and C. Kang (2013). Impact of wind power forecasting on unit commitment and dispatch. *IEEE Transactions on Power Systems*, 10.1109/TPWRS.2013.2254503.

Jiang, S. E. A. (1995). Polynomial loss models for economic dispatch and error estimation. *IEEE Transactions on Power Systems* 10(3), 1546–1552.

Kothari, D. and J. Dhillon (2004). *Power System Optimization*. New Delhi, India: Prentice-Hall of India.

Murty, P. (2009). *Operation and control in power systems*. Hyderabad, India: BS Publications.

Nanda, J., L. La, J. Ma, N. Rajkumar, A. Nanda, and M. Prasad (1999). Additional tests for a unit root allowing for a break in the trend function at an unknown time. *International Journal of Electrical Power and Energy Systems* 21, 555–560.

Wang, J., A. Botterud, M. Vladimiro, C. Monteiro, and G. Sheble (2009). Impact of wind power forecasting on unit commitment and dispatch. [www.dis.anl.gov/pubs/65610.pdf](http://www.dis.anl.gov/pubs/65610.pdf).

Zimmerman, R., C. Sanchez, and R. Thomas (2011). Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems* 26(1), 12–19.