

# DEVELOPMENT OF EQUIVALENT DYNAMIC MODEL OF DISTRIBUTION NETWORK USING CLUSTERING PROCEDURE

Samila Mat Zali

N. C. Woolley

J. V. Milanović

The University of Manchester, UK

[Samila.MatZali@postgrad.manchester.ac.uk](mailto:Samila.MatZali@postgrad.manchester.ac.uk), [Nick.Woolley@postgrad.manchester.ac.uk](mailto:Nick.Woolley@postgrad.manchester.ac.uk),  
[milanovic@manchester.ac.uk](mailto:milanovic@manchester.ac.uk)

**Abstract** – This paper presents the development of a dynamic equivalent model for a Distribution Network Cell (DNC) based on the UK 11kV distribution system. In a new and innovative approach, the methodology utilises the k-means clustering procedure to approximate the input and output dynamic responses of the network. By grouping together similar dynamic responses, the developed methodology is able to significantly reduce the computation time required to approximate the dynamic response of the network. The accuracy of the procedure is quantified by analysing the deviation of the true response from the approximated cluster centre. The computational efficiency of the approach is compared against parameter estimation without clustering, and the resulting approximation errors and computational improvements highlighted.

**Keywords:** *dynamic equivalent, grey-box model, parameter estimation, k-means, clustering*

## 1 INTRODUCTION

Increasingly high penetrations of renewable energy resources are renewing interest in modelling power system dynamic behaviour in both transmission and distribution networks. An equivalent dynamic model of a DNC is extremely important, as it enables power system operators to quickly estimate the impact of disturbances on power system dynamic behaviour. A dynamic equivalent works by simplifying the complexity of the distribution network and reducing the computation time required to run a full dynamic simulation. It offers a simple and low order representation of the system without comprising distribution network dynamic characteristics and behaviour as seen by the external grid.

The identification of dynamic equivalents can be a time consuming online (or offline) task. Ideally, the process of determining the dynamic equivalent model should be quick and accurate in order to help reduce overall computational time. This research proposes a new and novel method to reduce the computational time required to identify an online or offline dynamic equivalent model. The dynamic model is also accurate, estimating the dynamic response of a network as accurately as other more computationally intensive techniques.

## 2 DYNAMIC EQUIVALENT MODELLING

Previous research on dynamic equivalent models has been applied to both large power system networks[1] and MicroGrids (MGs) [2]. In [1], a grey-box approach

was applied to a large power system network and in [2] a grey-box approach was combined with an Evolutionary Particle Swarm Optimization (EPSO) algorithm for parameter estimation of an MG. The MG model used in this study focuses on representing the MG slow dynamics; similar to the active power control found in diesel engine run power generators. The model was developed in MATLAB using Simulink and the parameters estimated using an EPSO algorithm and the sum of error function used to measure the closeness of estimated model with the real system. The modelled MG only incorporates Solid Oxide Fuel Cell (SOFC) and high speed Single Shaft Microturbines (SSMT) (and their corresponding power electronic interfaces). Other types of distributed generation, for example, photovoltaics, wind generators and small hydro, were not considered in this research.

Some researchers have proposed developing dynamic equivalents of a DNC using a Hankel norm approximation [3-5] and a recurrent artificial neural network (ANN) [6, 7]. Another form of dynamic equivalent model of a DNC was developed using a system identification procedure and black-box modelling [8, 9]. The DNC was modelled as a black box due to a lack of detailed information on the network structure and parameters. The black box approach means that the dynamic equivalent of the distribution network is obtained based on observed input and output data. The voltage and frequency are used as the input, and real and reactive power as the output. The parameter identification is then performed by importing the input and output data into the MATLAB System Identification toolbox. The model is developed in the form of state space auto-regressive model with exogenous input (ARX). Its main advantage is its simplicity of implementation as it does not require detailed network information. The model however, is highly dependent on the type and location of the disturbance.

In this research a grey-box approach was chosen (as opposed to a black-box approach) because of its ability to incorporate prior knowledge about distribution network structure. This ensures that the developed model is physically relevant, intuitive and accurate. The equivalent model includes a converter-connected generator and a composite load model in parallel. The dynamic equivalent model is presented in the form of

sixth-order nonlinear state space equation and developed from the algebraic and differential equations describing typical components of a distribution network.

The goal of the grey box model is to estimate the dynamic response of real and reactive power based on the observation of voltage and frequency at the DNC connection point. All inputs (voltage and frequency) and outputs (real and reactive power) are measured at the connection point of the DNC following various disturbances. Recorded input and output signals are then used to estimate parameters of the nonlinear grey-box model using the parameter estimation procedure in MATLAB's System Identification Toolbox environment [19, 20].

The aim of this paper is to show how a clustering algorithm can increase the computational efficiency and reduce the model complexity of a dynamic equivalent model. In this research, the k-means clustering procedure is used to group together similar dynamic responses into a smaller set of approximate average responses. The parameter estimation procedure can then be performed on these average responses and the estimated parameters of a grey-box model can be obtained. The trade off between computational efficiency and accuracy of the dynamic equivalent model can then be explored by comparing the clustered grey-box model with the grey-box model obtained by estimating each individual input and output response.

### 3 GREY-BOX MODEL STRUCTURE

The equivalent model includes a converter-connected generator and a composite load model in parallel. The composite load model is a widely used and popular load model because it accurately models the physical aspects of the load and incorporates high proportions of induction motors [10-17]. High proportions of induction motors usually contribute significantly to a system's dynamic response as these motors typically consume 60 to 70% of the total energy supplied by a power system [18]. A composite load model was therefore chosen to represent the loads of DNC. Detailed development of the grey box model is presented in [19, 20] and only briefly discussed here.

The nonlinear state space model can be summarized as follows:

$$\begin{aligned} \dot{x} &= Ax + Bu + f(x) \\ y &= Cx + Du + f(x) \end{aligned} \quad (1)$$

where the A, B, C and D are the coefficient matrixes, x is the state vector, u is the input vector, y is the output vector and f(x) is a function that represents the nonlinear parts of the model. The grey-box model is a sixth order nonlinear state space model of the form shown in (1). Equation (2) is in general form where P1, P2, ..., P20 represent the unknown parameters [19, 20]. The model outputs, real power (P) and reactive power (Q),

are represented by the output vector y. The input vector u includes voltage (V) and frequency (f).

$$\begin{aligned} \dot{x} &= \begin{bmatrix} P1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ P2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & P3 & 0 & 0 \\ 0 & 0 & 0 & P4 & P5 & 0 \\ 0 & 0 & 0 & 0 & 0 & P6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} + \begin{bmatrix} P7 & 0 \\ P8 & -1 \\ 0 & 0 \\ P9 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ P10 \\ 0 \\ 0 \end{bmatrix} \\ y &= \begin{bmatrix} P11 & 0 & 0 & P12 & 0 & P13 \\ P14 & 0 & 0 & P15 & 0 & P16 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} + \begin{bmatrix} P17 & 0 \\ P18 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} P19 \\ P20 \end{bmatrix} \end{aligned} \quad (1)$$

### 4 PARAMETER ESTIMATION PROCEDURE

The parameter estimation procedure is shown in Figure 1. Initially, the input signals, voltage and frequency (V, f), are imported into the nonlinear grey-box model together with the initial values of model parameters to be estimated. The System Identification Toolbox in MATLAB is used to develop the grey-box model and to produce the output responses of real and reactive power ( $\hat{P}, \hat{Q}$ ). The simulated output responses ( $\hat{P}, \hat{Q}$ ) are obtained from the model using estimated model parameters based on nonlinear least square optimization according to the pre-specified identification criterion.

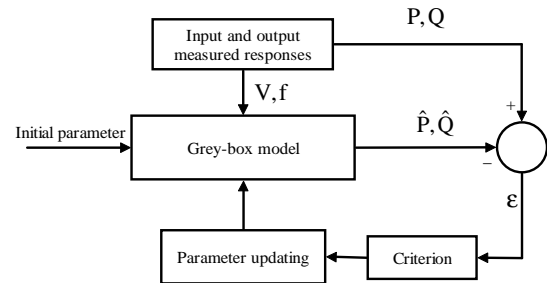


Figure 1: The parameter estimation procedure.

Nonlinear least square is a general technique used to fit a curve through data. It fits data to any equation that defines Y as a function of X and one or more parameters. It finds parameter values which generate a curve that come closest to the data (minimizes the sum of the squares of the vertical distances between data points and curve) [21]. This technique requires a model of the analysed signal. In this research, the signal model is defined by the grey-box model structure. The nonlinear least square optimization is generally used where the goal is to minimize the difference between the physical observation and the prediction from mathematical model. More precisely, the goal is to determine the best values of the unknown nonlinear parameters (P1 to P20) in order to minimize the squared errors between the measured values of the signal and the computed ones. The iterative process and parameter tuning continues until the error (ε) between responses

obtained from the model and actual, (measured or obtained through nonlinear simulations using full scale DNC model) output responses ( $P$  and  $Q$ ) is smaller than the predefined threshold.

## 5 TEST NETWORK

The distribution network study system is shown in Figure 2 and is broadly based on the UK 11 kV distribution network. The distribution network is connected to the 33 kV external grid, represented by equivalent synchronous generator source. The converter-connected generation (CCG) and fixed speed induction generators (FSIG) are connected on feeder 1 and doubly fed induction generators (DFIG) on feeder 3. Two synchronous generators (SG) connected to Bus 2, are driven by gas turbine units modelled as IEEE GAST type. Further details on distribution network modelling, in DigSILENT PowerFactory software, can be found in [22, 23].

In the following case studies, the total installed local generation (i.e. generation connected at Bus 1, see Figure 2) is equal to the total load in the DNC. Therefore there is no exchange of real power with the rest of distribution network through Bus 1. This condition was chosen so that the disturbances occurred in DNC have a minor effect on the transmission

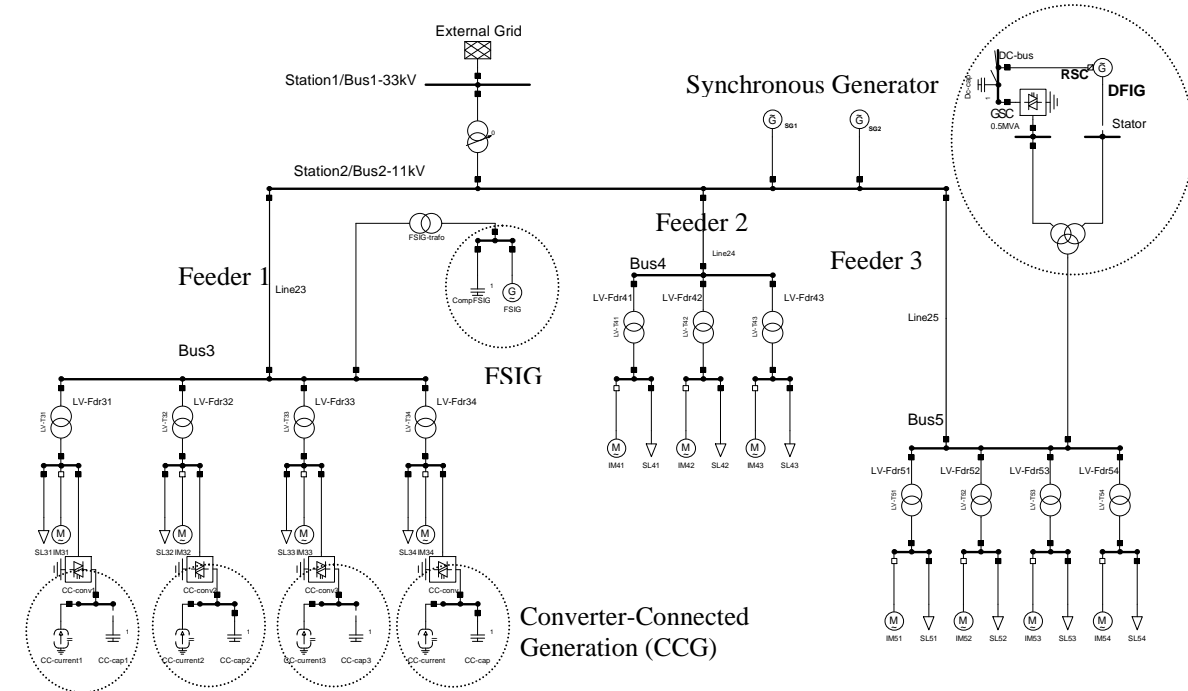
network. Thus the model derived from the measurements is the equivalent model of DNC system only.

The input and output responses due to a large disturbance (three-phase short circuit fault) are measured at Bus 1 (the grid connection point). Different case studies were considered in order to identify several possible dynamic responses of the DNC. The following figures, tables and associated discussion illustrate the results from a set of 34 simulation cases as described in Figure 2.

Each of the case studies was derived based on four static load models, two dynamic load compositions, five fault locations and seven DG composition scenarios as listed in Table 1. All of the responses are then used in the parameter estimation procedure to obtain a dynamic equivalent model. The initial parameter values for the estimation procedure are set using either typical values or parameters from published literature [10, 18, 24-26].

DG COMPOSITION	T1	T2	T3	T4	T5	T6	T7
SG (MW)	4.5	4.5	4.5	5.0	2.5	2.5	2.0
FSIG (MW)	3.0	0	0	0	0	0	0
DFIG (MW)	3.0	3.0	0	6.0	4.5	0	9.0
CCG (MW)	0.5	3.5	6.5	0	4.0	8.5	0

Table 1: Configuration parameters of the distributed generation.



### 34 Test Disturbance Scenarios

#### Static load model (across 5 locations, 20 studies)

- Constant impedance ( $Z$ )
- Constant current ( $I$ )
- Constant power ( $P$ )
- Composite load (ZIP)

#### Dynamic load composition (at 1 location, 2 studies)

- (1/3 small + 1/3 medium + 1/3 large) motors
- (2/3 small + 1/6 medium + 1/6 large) motors

#### Topologies (F 1 & F4 across 6 topologies, 12 studies)

- Topology T1 to T7

#### Fault location (fault location 1 & 4 across

- Fault at Bus 1 (F1)
- Fault at Bus 2 (F2)
- Fault at Bus 3 (F3)
- Fault at Bus 4 (F4)
- Fault at Bus 5 (F5)

Figure 2: The test network and the 34 test disturbance scenarios used in this research.

The disturbances are first simulated and the input and output responses were measured at the point of connection. All the responses are simulated for 9 s after the initial disturbance. The disturbance is simulated at 1 sec and then cleared after a further 500 ms. The sampling rate was 10 ms.

The performance of the developed equivalent model is evaluated by computing a best fit value. A best fit value is a percentage value that shows how well the simulated model output matches the measured output. The performance value is computed using (2) as below [27]:

$$\text{Best fit} = \left(1 - \frac{|y - \hat{y}|}{|y - \bar{y}|}\right) \times 100 \quad (2)$$

where  $y$  is the measured output,  $\hat{y}$  is the simulated model output and  $\bar{y}$  is the mean of  $y$ . 100% corresponds to a perfect fit while 0% indicates that the fit is no better than guessing the output to be a constant ( $\hat{y} = \bar{y}$ ).

## 6 CLUSTERING PROCEDURE

### 6.1 *k*-means Clustering Algorithm

The *k*-means algorithm is a clustering method which aims to partition a dataset of  $n$  observations into  $k$  clusters [28]. For example, for a set of observations  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$  (where each observation is a  $d$  dimensional vector) the aim of *k*-means is to assign each observation to a cluster, whilst minimising the within cluster sum of squares, as shown in (3).

$$\arg \min_K \sum_{i=1}^k \sum_{\mathbf{x}_j \in K} \|\mathbf{x}_j - \mu_i\| \quad (3)$$

where  $K$  is the set of  $d$  dimensional cluster centres, and  $\mu_i$  is the mean of the points in the  $i$  th cluster centre. The algorithmic implementation for *k*-means used in this paper is available in the MATLAB programming environment.

The *k*-means clustering procedure was used because it is conceptually simple, widely implemented and is probably the most popular clustering procedure used by researchers [29].

### 6.2 Clustering Power Traces

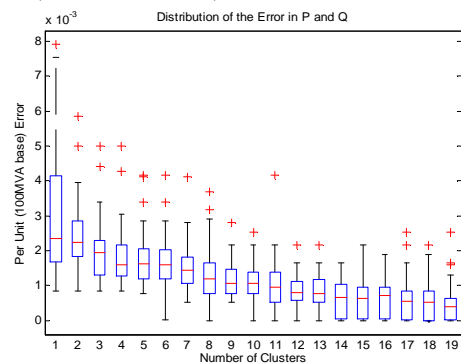
In this research, the real power and reactive power responses of the network at bus 1 were used as inputs to the *k*-means clustering algorithm. The active power and reactive power values at bus 1 were re-sampled across the 10 s trace window into 10 ms increments to form two observation vectors with 1000 measurements. Each fault study created a set of power, reactive power, voltage and frequency traces across the 10 s time interval at bus 1. Each of these reactive power ( $\mathbf{q}_i \in P$ ) and active power vectors ( $\mathbf{p}_i \in P$ ) were then combined into one single vector ( $\mathbf{x}_i \in X$ ) with 2000 elements.

The *k*-means algorithm was then applied to the set  $X$  to partition the set into a group of  $k$  clusters. Each grouping can be labelled with a centroid [28]. The cluster centroid defines a representative location in  $d$  dimensional space for all the members of that particular cluster. This location can be used as an approximation for all of the values in the cluster.

Selecting the number of clusters to divide the data set into is a trade off between the error in the approximation and the speed improvement of the dynamic simulation gained by reducing the number of clusters. A greater number of clusters results in a more accurate classification of the dynamic responses. In this research, 4 clusters were selected to group the 34 scenarios. This number was chosen by analysing the graphical grouping of the clusters (as shown in Figure 3) to ensure that obvious differences in the traces were incorporated into each approximation. The distribution of the average absolute difference in both reactive and active power can also be used as a guide (as shown in equation (4)) to select the number of clusters.

$$e_j(\mathbf{x}_i) = \sum_{m=1}^d \frac{|x_{j,m} - k_{j,m}|}{d} \quad (4)$$

Where  $x_{j,m}$  is the  $m$  th element of the  $d$  dimensional vector representing real and reactive power in the  $j$  th cluster, and  $e_j(\mathbf{x}_j)$  is a scalar measurement of the accuracy of the  $\mathbf{x}_i$  fault scenario. The distribution of the per unit error ( $e_j(\mathbf{x}_j)$ ) across various numbers of clusters is shown in Figure 3. Figure 3 shows that with one cluster, the median distance across all fault scenarios to a single centroid found with *k*-means is 0.0023 per unit. The error  $e_j(\mathbf{x}_j)$  shown in Figure 3 will decline to zero if 34 clusters were selected with *k*-means, as this would represent an individual centroid for each of the 34 fault scenarios. Figure 3 shows that 4 is a reasonable choice for the number of clusters as 3 clusters has a median average error 67% larger, with little change between the median error reduction with 5 clusters (or more clusters).



**Figure 3:** The distribution of the average per unit error across various numbers of clusters obtained with *k*-means.

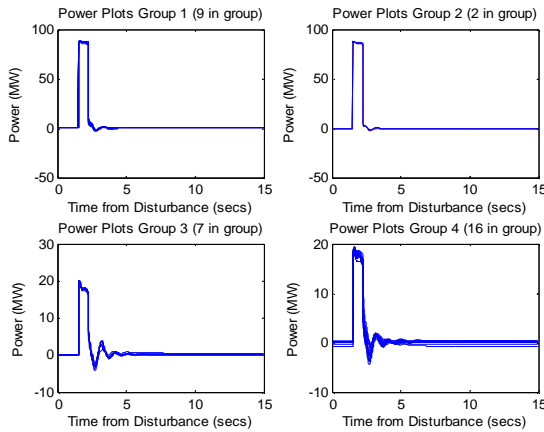
Figure 4 shows the approximated power traces and the real power traces when they are grouped into four clusters using *k*-means.

## 7 RESULTS

The parameter estimation procedure was performed on both the 34 fault case studies and the 4 approximate responses produced as a result of clustering. Each of the 34 grey-box models was compared against its corresponding grey-box cluster model and also against the true dynamic response. Ideally, the grey-box model for an individual disturbance and the grey box model for a cluster centroid should closely match the true dynamic response of the system.

### 7.1 Parameter Estimation for 34 Disturbances

Table 2 shows the results of grey box model fitting for each of the 34 disturbances, compared against the true dynamic response. The ranges of the best fit percentages are segmented by fault location, real power (P) and reactive power (Q). The table shows that by estimating parameters for each of the 34 disturbances, the best fit percentages (equation (2)) range from 80% to 94%.



**Figure 4:** The approximate and real power traces achieved when grouping the power and reactive power traces of 34 different scenarios into four clusters using k-means.

Table 2 shows how the clustering procedure partitions each of the 34 different fault scenarios.

FAULT LOCATION	TOPOLOGY	LOAD MODEL	ASSIGNED CLUSTER
1	Topology 1	I, P, Z & ZIP	1
2	Topology 1	I, P, Z & ZIP	3
3	Topology 1	I, P, Z & ZIP	4
4	Topology 1	I, P, Z & ZIP	4
5	Topology 1	I, P & ZIP	3
5	Topology 1	Z	4
1	Topology 2, 3, 5 & 6	ZIP	1
1	Topology 4 & 7	ZIP	2
4	Topology 2, 3, 4, 5, 6 & 7	ZIP	4
Motor 1 (FL 2)	Topology 1	Dynamic	1
Motor 2 (FL 2)	Topology 1	Dynamic	4

**Table 2:** Cluster groups of the 34 simulated fault scenarios.

CASES	BEST FIT (%)	
	P RESPONSE	Q RESPONSE
Fault at Bus 1 (F1)	86.84 – 89.25	90.34 – 92.70
Fault at Bus 2 (F2)	80.44 – 82.46	90.27 – 93.56
Fault at Bus 3 (F3)	80.92 – 83.01	90.91 – 92.54
Fault at Bus 4 (F4)	80.13 – 81.09	90.20 – 91.99
Fault at Bus 5 (F5)	81.05 – 84.35	92.05 – 93.40

**Table 3:** Best fit value of simulated response by grey-box model

Table 4 shows the range of grey box estimated model parameter values segmented according to fault location. Table 4 shows the range of estimated model parameters obtained when performing parameter estimation on each of the 34 disturbance scenarios.

CASES	PARAMETERS									
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Fault at Bus 1 (F1)	-39.826 to -30.869	0.025	-3.913 to -3.480	-0.080	-0.984	-10.560	-2.534 to 2.336	-10.138	-14.263 to -8.575	8.610 to 13.921
	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Fault at Bus 2 (F2)	29.812 to 33.704	-43.365 to -39.008	0.392	-9.150 to -4.467	11.839 to 16.281	0.392	-281.598 to -208.288	-358.871 to -345.756	208.059 to 281.092	351.073 to 363.650
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Fault at Bus 3 (F3)	-40.544 to -19.046	0.025	-4.913 to -3.823	-0.080	-0.984	-10.560	-4.413 to -1.681	-10.138	-17.990 to -15.848	15.459 to 17.518
	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Fault at Bus 4 (F4)	29.743 to 32.881	-45.787 to -42.973	0.392	-6.399 to -2.710	9.436 to 11.496	0.392	-313.363 to -293.902	-365.714 to -355.594	293.949 to 313.474	360.863 to 371.067
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Fault at Bus 5 (F5)	-37.091 to -20.376	0.025	-7.473 to -3.591	-0.080	-0.984	-10.560	-7.495 to -1.406	-10.138	-19.507 to -12.221	12.047 to 18.351
	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Fault at Bus 5 (F5)	30.658 to 33.286	-46.947 to -42.218	0.392	-9.109 to -3.614	7.062 to 11.732	0.392	-296.850 to -261.575	-370.992 to -355.595	261.710 to 297.833	361.327 to 375.128
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Fault at Bus 5 (F5)	-36.012 to -21.934	0.025	-3.965 to -3.220	-0.080	-0.984	-10.560	-3.286 to 2.656	-10.138	-15.355 to -9.532	9.663 to 14.961
	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Fault at Bus 5 (F5)	26.299 to 32.264	-45.193 to -41.861	0.392	-22.182 to 0.726	7.687 to 9.019	0.392	-284.205 to -232.718	-365.052 to -360.459	232.799 to 285.240	365.500 to 370.743

**Table 4:** The estimated parameter values for each of the 20 grey-box model parameters obtained when estimating parameters of each of the 34 disturbances individually.

CASES	PARAMETERS									
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Cluster 1	-37.903	0.025	-5.430	-0.080	-0.984	-10.560	-7.078	-10.138	-5.882	4.263
Cluster 2	-38.492	0.025	-6.016	-0.080	-0.984	-10.560	-7.455	-10.138	-5.551	4.713
	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Cluster 1	30.669	-34.422	0.392	-18.791	25.037	0.392	-231.659	191.627	233.030	196.463
Cluster 2	27.398	-33.249	0.392	-19.241	24.432	0.392	-231.359	-190.640	230.948	195.578
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Cluster 3	-36.400	0.025	-5.080	-0.080	-0.984	-10.560	-2.947	-10.138	-12.980	12.754
Cluster 4	-29.930	0.025	-5.683	-0.080	-0.984	-10.560	-1.984	-10.138	-18.999	17.952
	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Cluster 3	30.474	-41.008	0.392	-4.764	16.484	0.392	-223.544	-356.185	225.054	361.141
Cluster 4	32.584	-47.024	0.392	-5.548	9.784	0.392	-294.120	-357.664	294.516	362.709

Table 5: The estimated parameter values for each of the 20 grey-box model parameters obtained when estimating parameters using the 4 cluster centroids.

### 7.2 Parameter Estimation for 4 Cluster Centroids

Table 5 shows the grey-box model parameter values for each of the 4 cluster centroids. A quick comparison of the parameter values in Table 5 with those presented in Table 4 shows that the grey box parameter values for the cluster centroids are within the parameter ranges for the individual cluster members. For example, cluster 4 represents all faults at location 5 (see Table 2) and has a P1 value of -29.9 which is within the range reported in Table 4 of -21.9 to -36.0.

Table 6 presents the best fit values for the estimated model parameters of each cluster compared against the true dynamic response. The best fit percentages range from 81% to 93% as opposed to 80% to 94% when the parameters are estimated for each of the 34 disturbances individually (Table 3).

CLUSTER	BEST FIT (%)	
	P RESPONSE	Q RESPONSE
1	88.17	92.91
2	88.28	90.98
3	80.67	89.13
4	81.59	90.74

Table 6: Best fit value of estimated model for clusters 1 to 4.

### 7.3 Graphical Comparison of Results

Figure 5 to Figure 7 shows three comparisons of the real and reactive power dynamic response for four specific disturbances. The graphs compare the actual dynamic response of the network (straight line) against an estimated model built for a specific disturbance (dotted line) and an estimated model built using a cluster centroid (dashed line).

Figures 5 to 8 show that the grey-box models visually approximate the network's dynamic response to a high level of accuracy. There is no discernable difference between a disturbance specific grey-box model and a grey-box model based on cluster centroids. This corroborates with the results in Table 3 and Table 6 which showed little difference in the estimation accuracy of a cluster centroid or disturbance specific grey-box model.

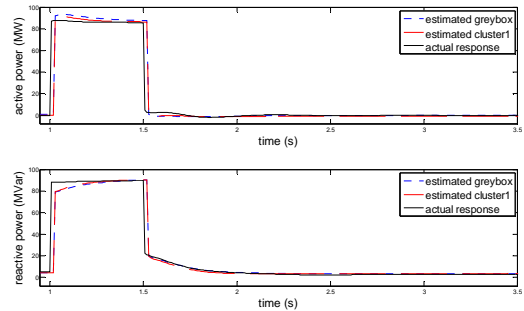


Figure 5: The real and reactive power response of a disturbance at fault location 1, topology 1 and ZIP load model compared against an estimated response for this specific disturbance and the estimated response obtained from cluster 1.

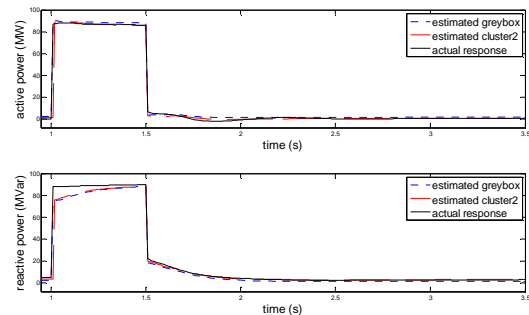


Figure 6: The real and reactive power response of a disturbance at fault location 1, topology 4 and ZIP load model compared against an estimated response for this specific disturbance and the estimated response obtained from cluster 2.

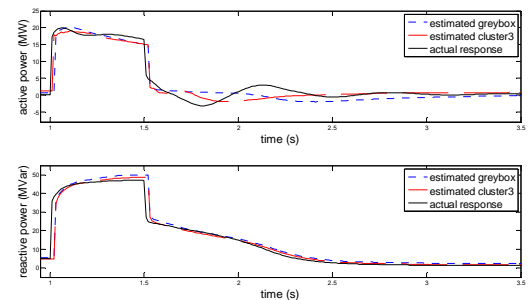
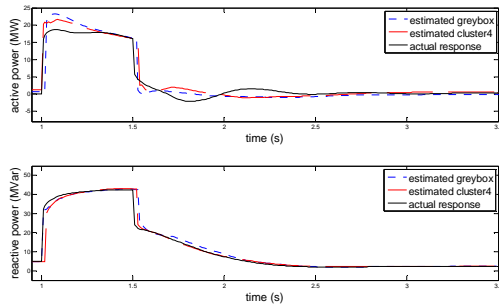


Figure 7: The real and reactive power response of a disturbance at fault location 5, topology 1 with constant P load model compared against an estimated response for this specific disturbance and the estimated response obtained from cluster 3.

## 8 CONCLUSION

This paper introduced a k-means clustering process into the grey-box model parameter value estimation procedure to improve the computational performance of parameter value estimation. The proposed clustering method was tested using a series of disturbance scenarios, and shown to be indiscernible from the standard grey box parameter value estimation procedure, whilst still closely approximating the dynamic response of the network. The technique is viable and robust method to calculate the dynamic equivalent model of a distribution network cell (DNC) using case studies built from a wide variety of network fault and composition scenarios.



**Figure 8:** The real and reactive power response of a disturbance at fault location 4, topology 1 and a motor load model compared against an estimated response for this specific disturbance and the estimated response obtained from cluster 4.

### 7.4 Computational Performance

Clustering the dynamic responses before performing parameter estimation reduces the computational complexity of grey-box modelling. It is therefore important to quantify the increase in computational performance that can be achieved with the technique.

Estimating the grey-box parameter values for a disturbance specific grey-box model took 2 minutes to compute using the MATLAB System Identification toolbox. Estimating the parameter values for a grey-box model based on a cluster-centroid takes the same length of time. Clustering the dynamic responses takes roughly 10 ms. Performing clustering before parameter value estimation therefore reduced this simulation by 60 minutes. This represents a reduction in computational time of 88%.

It should be noted that calculations were carried out on a Pentium 4 2.8 GHz workstation, running Matlab 7.7. Both the clustering and parameter value estimation were carried out using Matlab's internal functions.

### 7.5 Analysis of the Results

Tables 2 to 5 and figures 5 to 8 clearly show that the estimated cluster centroids and disturbance specific grey-box models are nearly identical in terms of estimation accuracy. It was hypothesised that the clustering procedure would be as accurate as disturbance specific parameter estimation; the results do seem to support this assertion. The accuracy of the technique is maintained by intelligently grouping together similar dynamic responses, rather than changing the grey-box model estimation procedure. Indeed, tables 3 and 4 show that the parameter values obtained using the cluster centroid estimation procedure reside within the same range as those parameters found for a disturbance specific grey box model.

By including the clustering procedure in the parameter estimation process, the overall computational time of the dynamic equivalent determination task has been significantly reduced. Clustering the data before parameter estimation ensures that parameter estimation is run once for four specific cases, rather than 34 times for each specific disturbance. The number of parameter value estimations has been reduced by about 88%.

## REFERENCES

- [1] A. M. Stankovic and A. T. Saric, "Transient power system analysis with measurement-based gray box and hybrid dynamic equivalents," *Power Systems, IEEE Transactions on*, vol. 19, pp. 455-462, 2004.
- [2] F. O. Resende and J. A. Pecas Lopes, "Development of Dynamic Equivalents for MicroGrids using System Identification Theory," in *Power Tech, 2007 IEEE Lausanne, 2007*, pp. 1033-1038.
- [3] A. Ishchenko, A. Jokic, J. M. A. Myrzik, and W. L. Kling, "Dynamic reduction of distribution networks with dispersed generation," in *Future Power Systems, 2005 International Conference on*, 2005, pp. 7 pp.-7.
- [4] A. Ishchenko, J. M. A. Myrzik, and W. L. Kling, "Dynamic equivalencing of distribution networks with dispersed generation using Hankel norm approximation," *Generation, Transmission & Distribution, IET*, vol. 1, pp. 818-825, 2007.
- [5] A. Ischchenko, J. M. A. Myrzik, and W. L. Kling, "Dynamic equivalencing of distribution networks with dispersed generation," in *Power Engineering Society General Meeting, 2006. IEEE, 2006*, p. 8 pp.
- [6] A. M. Azmy and I. Erlich, "Identification of dynamic equivalents for distribution power networks using recurrent ANNs," in *Power Systems Conference and Exposition, 2004. IEEE PES, 2004*, pp. 348-353 vol.1.
- [7] A. M. Azmy, I. Erlich, and P. Sowa, "Artificial neural network-based dynamic equivalents for distribution systems containing active sources," *Generation, Transmission and Distribution, IEE Proceedings-*, vol. 151, pp. 681-688, 2004.
- [8] X. Feng, Z. Lubosny, and J. Bialek, "Dynamic Equivalencing of Distribution Network with High Penetration of Distributed Generation," in *Universities Power Engineering Conference, 2006. UPEC '06. Proceedings of the 41st International, 2006*, pp. 467-471.
- [9] X. Feng, Z. Lubosny, and J. W. Bialek, "Identification based Dynamic Equivalencing," in *Power Tech, 2007 IEEE Lausanne, 2007*, pp. 267-272.
- [10] C. Byoung-Kon, C. Hsiao-Dong, L. Yinhong, C. Yung-Tien, H. Der-Hua, and M. G. Lauby, "Development of composite load models of power systems using on-line measurement data," in *Power Engineering Society General Meeting, 2006. IEEE, 2006*, p. 8 pp.

- [11] H.-D. Chiang, J.-C. Wang, C.-T. Huang, Y.-T. Chen, and C.-H. Huang, "Development of a dynamic ZIP-motor load model from on-line field measurements," *International Journal of Electrical Power & Energy Systems*, vol. 19, pp. 459-468, 1997.
- [12] R.-M. He, J.-L. Wang, J. Ma, Y.-H. Xu, and D. Han, "Impacts of DFIG-based wind farm on load modeling," in *Power & Energy Society General Meeting, 2009. PES '09. IEEE*, 2009, pp. 1-6.
- [13] M. Jin, H. Dong, H. Ren-Mu, D. Zhao-Yang, and D. J. Hill, "Reducing Identified Parameters of Measurement-Based Composite Load Model," *Power Systems, IEEE Transactions on*, vol. 23, pp. 76-83, 2008.
- [14] M. Jin, H. Renmu, and D. J. Hill, "Composite load modeling via measurement approach," in *Power Engineering Society General Meeting, 2006. IEEE*, 2006, p. 1 pp.
- [15] P. Ju, F. Wu, Z. Y. Shao, X. P. Zhang, H. J. Fu, P. F. Zhang, N. Q. He, and J. D. Han, "Composite load models based on field measurements and their applications in dynamic analysis," *Generation, Transmission & Distribution, IET*, vol. 1, pp. 724-730, 2007.
- [16] H. Renmu, J. Ma, and D. J. Hill, "Composite load modeling via measurement approach," *Power Systems, IEEE Transactions on*, vol. 21, pp. 663-672, 2006.
- [17] J. Wang, R. He, and J. Ma, "Load Modeling Considering Distributed Generation," in *Power Tech, 2007 IEEE Lausanne*, 2007, pp. 1072-1077.
- [18] P. Kundur, "Power System Stability and Control," *McGraw-Hill*, 1994.
- [19] S. M. Zali and J. V. Milanovic, "Modelling of Distribution Network Cell Based on Grey-box Approach," in *Med Power 2010*, 2010.
- [20] S. M. Zali and J. V. Milanovic, "Validation of Developed Grey-box Model of Distribution Network Cell," in *MedPower 2010*, 2010.
- [21] W. Sun and Y.-x. Yuan, *Optimization Theory and Methods: Nonlinear Programming* Springer-Verlag New York Inc, 2006.
- [22] J. V. Milanović and M. Kayikci, "Transient Responses of Distribution Network Cell with Renewable Generation," in *CD ROM of the IEEE Power System Conferences and Exposition, PSCE'06*, Atlanta, Georgia, USA, 2006.
- [23] M. Kayikchi, "The influence of wind plant control on transient performance of the network," *PhD thesis, University of Manchester*, 2007.
- [24] M. Dehghani and S. K. Y. Nikraves, "Nonlinear state space model identification of synchronous generators," *Electric Power Systems Research*, vol. 78, pp. 926-940, 2008.
- [25] W. Feng, Z. Xiao-Ping, and J. Ping, "Impact of wind turbines on power system stability," in *Bulk Power System Dynamics and Control - VII. Revitalizing Operational Reliability, 2007 iREP Symposium*, 2007, pp. 1-7.
- [26] W. Feng, Z. Xiao-Ping, and J. Ping, "Modeling and control of the wind turbine with the Direct Drive Permanent Magnet Generator integrated to power grid," in *Electric Utility Deregulation and Restructuring and Power Technologies, 2008. DRPT 2008. Third International Conference on*, 2008, pp. 57-60.
- [27] L. Ljung, *System Identification Toolbox™ 7 User's Guide: The Mathworks, Inc*, 2010.
- [28] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," in *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281-297.
- [29] D. T. Pham, S. Otri, A. Afify, M. Mahmaddin, and H. Al-Jabbouli, "Data Clustering Using the Bees Algorithm," in *40th CIRP International Manufacturing Systems Seminar*, 2007.