Hybrid Intelligent Model to Predict the SOC of a LFP Power Cell type

Luis Alfonso Fernández-Serantes[1], Raúl Estrada Vázquez[2], José Luis Casteleiro-Roca[1], José Luis Calvo-Rolle[1], and Emilio Corchado[3]

[1] Universidad de A Coruña,
Departamento de Ingeniería Industrial,
Avda. 19 de febrero s/n, 15495, Ferrol, A Coruña, España
luis.alfonso.fernandez.serantes@udc.es

[2] FH JOANNEUM University of Applied Sciences,
Institute for Electronic Engineering,
Werk-VI-Strasse 46, 8605, Kapfenberg, Austria

[3] Universidad de Salamanca,
Departamento de Informática y Automática,
Plaza de la Merced s/n, 37008, Salamanca, Salamanca, España

Abstract. Nowadays, batteries have two main purposes: to enable mobility and to buffer intermitent power generation facilities. Due to their electromechaminal nature, several tests are made to check battery performance, and it is very helpful to know a priori how it works in each case. Batteries, in general terms, have a complex behavior. This study describes a hybrid intelligent model aimed to predict the State Of Charge of a LFP (Lithium Iron Phosphate - LiFePO4) power cell type, deploying the results of a Capacity Confirmation Test of a battery. A large set of operating points is obtained from a real system to create the dataset for the operation range of the power cell. Clusters of the different behavior zones have been obtained to achieve the final solution. Several simple regression methods have been carried out for each cluster. Polynomial Regression, Artificial Neural Networks and Ensemble Regression were the combined techniques to develop the hybrid intelligent model proposed. The novel model allows achieving good results in all the operating range.

Keywords: Power cell, battery, clustering, neural networks

1 Introduction

The variation of the energy production at renewable energy installations like wind farms, the necessity of finding a substitute for fossil fuels in vehicles or to supply energy to portable devices, are several reasons for which electric energy storage is one of the trending solutions [1].

At present, there are some researches with the aim of improving energy storage. For instance, Smart Grid is a good example where energy store systems are employed to face intermittent renewable generations [2], such as wind or solar power production. In this way, portable devices require higher autonomy and lower weight with the purpose of improving people's quality of life [3].

In other terms, the current development of electric vehicles possesses the problem of storing energy, despite the fact that electric powertrains are more efficient than internal combustion engines [4]. Among the different energy storage technologies, this paper is focused on one of these technologies, the battery storage systems, specifically LFP (Lithium Iron Phosphate - LiFePO4) power cell type. Due to the relevance of this types of batteries, modeling them is really important, especially its behavior and its ageing prediction, as charging and discharging cycles reduce cell efficiency [5].

The classic regression models are based on Multiple Regression Analysis (MRA) methods [6]. MRA-based methods are useful due to their applications in different subjects [7, 8]; the first cite shows a model for cost prediction in the early state of projects, the second one proposes a method to evaluate suppliers performance. The main problem of these methods is their limitations in certain cases. For instances in [6] and [9] the common trouble is its non linearity and the different ways followed to solve them with aproaches based on MRA techniques. Regression techniques based on Soft Computing could avoid some of the problems mentioned above. Several works have been developed with this goal. In [10] the prediction state of a model predictive control system is carried out by meta-classifiers. By combining multi regression analysis and artificial neural networks an optimizing overbreak prediction is made in [11]. In [12] failure detection and prediction in wind turbines is achieved by using intelligent techniques.

Despite the new methods to solve regression problems, there are cases where it is not possible to achieve a good performance of the model, for instances due to the high non-linearity of the system. Clustering could be a complementary solution as a previous step to apply regression to the dataset [13]. K-means clustering algorithms are often employed with this purpose [14, 15]. With this method, all the dataset is divided into subsets (clusters), depending on the features of the input data. Then, regression is made over each cluster. Previous works like [16, 17] used similar techniques to solve other physical systems.

This study implements a hybrid model to predict several parameters in one specific test of batteries. To develop the model, K-means clustering algorithm was used to make groups of data with the same behavior. Then, three different regression techniques were tested for each group to choose the best one based on the lowest mean squared error achieved.

This paper is organized in the following way. After this introduction, the case of study section describes the employed test and how the dataset was obtained. Then, the model approach and the tested algorithms taken into account in the research are presented. The results section shows the best configuration achieved by the hybrid model. After the results, the conclusions and future works are presented.

2 Case of study

The model has been obtained to study the behavior of a LFP power cell type, by detecting its State Of Charge (SOC). The scheme of the practical implementation to carry out the test is shown on figure 1.

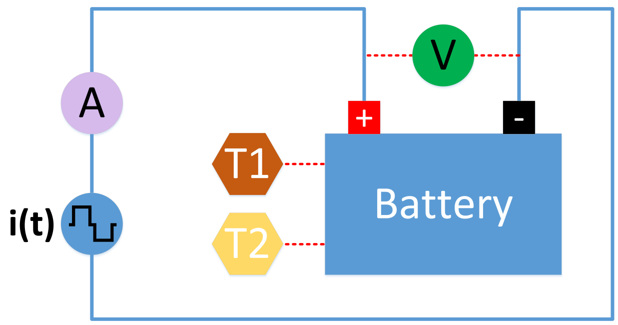

Fig. 1. Scheme of the capacity confirmation test

In the next subsections the test and the battery device are explained in detail.

2.1 The battery

A battery is a device capable of storing electricity within a electrochemical medium and reconverting it to electrical energy by electrochemical reactions [1]. The operating principle is based on a redox reaction by reducing the cations at the cathode and oxiding the anions at the anode, during the discharge, and in the other way during the charge [1]. This cycle can be repeated for a certain number of times, after that capacity decreases to anymore usable levels [18].

Lithium-ion (Li-ion) cells are one type of rechargeable batteries that have traditionally been used to power consumer electronic devices, and more recently for electric vehicles [19]. These cells are characterized for their light weight and high energy densities [1], they also have no memory effect, long life cycle and a low selfdischarge [19, 18].

2.2 Capacity confirmation of the battery test

The developed test measures the device capacity in ampere-hour at a constant current [20]. The first step is to charge the cell to its maximum SOC. After that, the battery is discharged at constant current up to the discharge voltage limit specified by the manufacturer [20]. Once the cell is recharged to its maximum capacity, the battery capacity and the SOC are calculated at each moment.

The test was done with a battery tester that can charge and discharge the cells at constant current, and it is able to measure different parameters. These parameters are the voltage provided by the battery, the current flowing to and from the battery, its temperature and the time, while the test is running.

The test scheme is shown on figure 1. On it, it is possible to see different components like a voltmeter *(V)*, an amperemeter *(A)*, and two temperature sensors *(T1 & T2)* to measure the temperature value at two different places. Also, there is a current source that provides and absorbs the flowing current *(i(t))*.

The cell used during this test was the LiFeBATT X-1P [21]. This power cell is a Lithium Iron Phosphate - LiFePO4, whose nominal capacity is $8000mAh$ and its nominal voltage $3.3V$. During the test (shown in figure 2), the next steps are carried out:

- (1) Charge: where the voltage increases from $3V$ to $3.65V$.
- (2) Rest after a charging process: where the voltage decreases up to the nominal value of $3.3V$.
- (3) Discharge: where the voltage decrease from $3.3V$ to $2V$.
- (4) Rest after discharging process: where the voltage grows up to the value of $3V$, and then the cycle starts again.

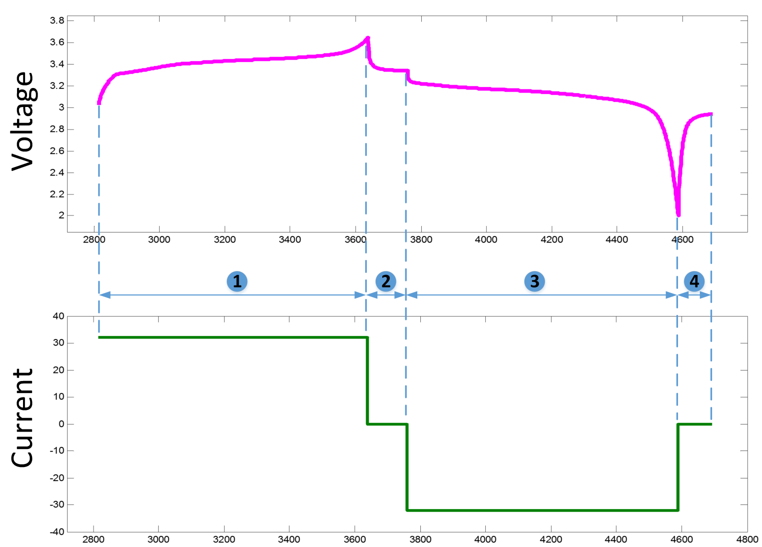

Fig. 2. Voltage and current during one cycle test

The analysis of voltage progress for one entire cycle is shown at the top of figure 2. The analysis of the current (bottom of figure 2) shows that the process carried out was done at a constant value of current. The current is positive when it flows from the source to the battery, and it is negative when it flows from the battery to the source.

With the value of current at each time it is possible to obtain the energy provided or absorbed in ampere-hour. If this energy is represented (top of figure 3) it is possible to see how the battery SOC increases during the charging period till 100% of charge. On the other hand, the SOC of the cell decreases till its minimum value of 0% during the discharging process.

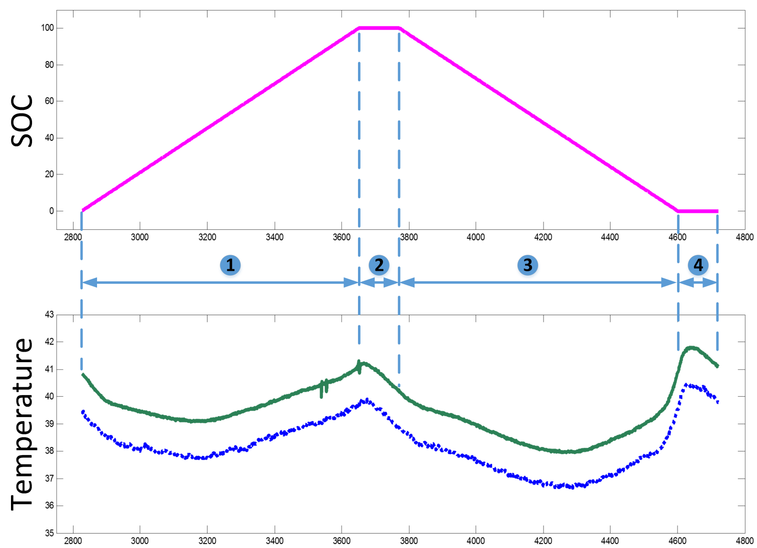

Fig. 3. Energy balance and temperatures during one cycle test

The measurement of temperatures are done with two sensors located at different places of the battery. These parameters vary cyclically depending on the state of the battery (charge, discharge, rest after charge and rest after discharge) and on its voltage. At the bottom of figure 3, it is possible to see the temperature behavior for each operating region.

The dataset has been obtained by carrying out the mentioned test over the power cell. The current and the voltage were registered to study the state of the battery. The parameter SOC was calculated with the current and the time for each test. Also two different temperatures were measured to detect malfunction on the device, if the temperature is far from the predicted one. The data were labeled during the test to know the corresponding state.

3 Model approach

The scheme of the model approach is shown in figure 4. Taking into account the power cell performance and the test made, it is possible to divide the dataset in four operation ranges. Consequently, four clusters are created and, three regression models (one per output) are implemented for each one. As shown on the figure 4, the global model has two inputs (current and voltage) and three outputs (SOC, T1 and T2). The cluster selector block connects the chosen models with the output.

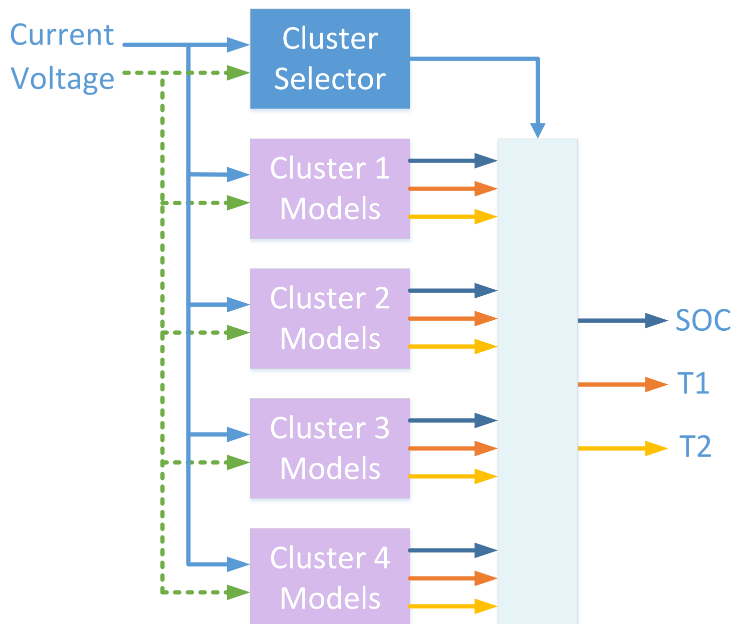

Fig. 4. Model approach

3.1 Techniques used

The techniques tested in the study to achieve the best model are described below.

Data Clustering. The K-means algorithm. Clustering is an unsupervised technique of data grouping where similarity is measured [22, 23]. Clustering algorithms try to organize unlabeled feature vectors into clusters or groups, in such a way that samples within a cluster are similar to each other [24]. K-means algorithm is a commonly used partitional clustering algorithm with square-error criterion, which minimizes error function shows in equation 1.

$$e = \sum_{k=1}^{C} \sum_{x \epsilon Q_k} \|x - c_k\|^2 \tag{1}$$

The final clustering will depend on the initial cluster centroids and on the value of K (number of clusters). Choosing K value is the most critical election because it requires certain prior knowledge of the number of clusters present in the data, which is highly doubtful. The K-means partitional clustering algorithm is computationally effective and works well if the data are close to its cluster, and the cluster is hyperspherical in shape and well-separated in the hyperspace.

Polynomial regression. Generally, a polynomial regression model [25] may also be defined as a linear summation of basis functions. The number of basis functions depends on the number of the model inputs, and the degree of the polynomial used.

With a degree 1, the linear summation could be defined as the one shown in equation 2. The model becomes more complex as the degree increases, equation 3 shows a second polynomial degree for the model.

$$F(x) = a_0 + a_1 x_1 + a_2 x_2 \tag{2}$$

$$F(x) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + a_4 x_1^2 + a_5 x_2^2 \tag{3}$$

Artificial Neural Networks (ANN): MultiLayer Perceptron (MLP). A multilayer perceptron is a feedforward artificial neural network [25]. It is one of the most typical ANNs due to its robustness and relatively simple structure. However, the ANN architecture must be well selected to obtain good results.

The MLP is composed by one input layer, one or more hidden layers and one output layer, all of them made of neurons and with pondered connections between the neurons of each layer.

Ensemble regression. The ensembles are a learning method usually employed for classification tasks [26]. Furthermore, this technique can be used for regression purposes with very satisfactory results when the dataset is large [27]. Regularization is a process for choosing fewer weak learners for an ensemble with the aim to increase predictive performance. Then it is possible to regularize regression ensembles. The method tries to find a set of weights α_t that minimize the expresion 4.

$$\sum_{n=1}^{N} w_n g\left(\left(\sum_{t=1}^{T} \alpha_t h_t(x_n)\right), y_n\right) + \lambda \sum_{t=1}^{T} |\alpha_t| \tag{4}$$

where,

- $\lambda \geq 0$ is the regularization parameter.
- h_t is a weak learner in the ensemble trained on N observations with predictors x_n, responses y_n, and weights w_n.
- $g(f, y) = (f - y)^2$ is the square error.

3.2 Preprocessing the dataset

The SOC of the battery should be from 0% to 100% and from 100% to 0%. In order to achieve this fact, some incomplete cycles were discarded. Taken this fact into account, given that the whole data acquisition has twelve cycles, only nine cycles were included to calculate the model. The data were recorded with a sample time of one second, and with the explained discard, the dataset was reduced from 18130 to 16369 samples.

The first technique applied to the dataset was clustering. To do that, K-means algorithm was applied and four clusters were created. These groups represent the different states of the cell test. In figure 5, it is possible to distinguish these four clusters: blue data correspond to state (1), magenta to state (2), red to (3) and green to (4).

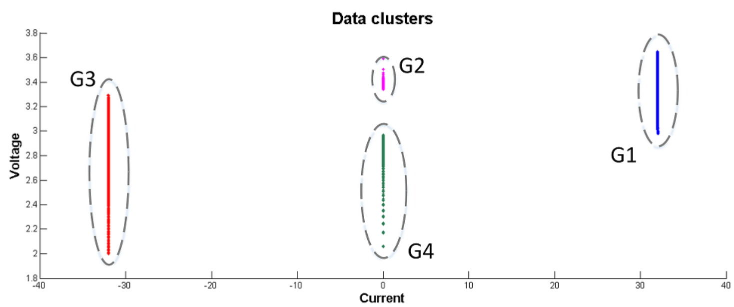

Fig. 5. Dataset clusters

All the dataset was divided in two parts to train and test the models. After this separation, each data was clustered. Table 1 shows the different number of samples for each cluster.

Cluster	Training	Testing	Total
(1) - Charge	4376	2243	6619
(2) - Rest	727	362	1089
(3) - Discharge	4975	2476	7451
(4) - Rest	835	375	1210
Total	10913	5456	16369

Table 1. Samples assigned to train and test the models

The three mentioned regression techniques were trained for the four clusters, one by each output of the model. As an example, the figure 6 shows the temperature in sensor 2. The four colors indicate the different clusters, as it was mentioned above.

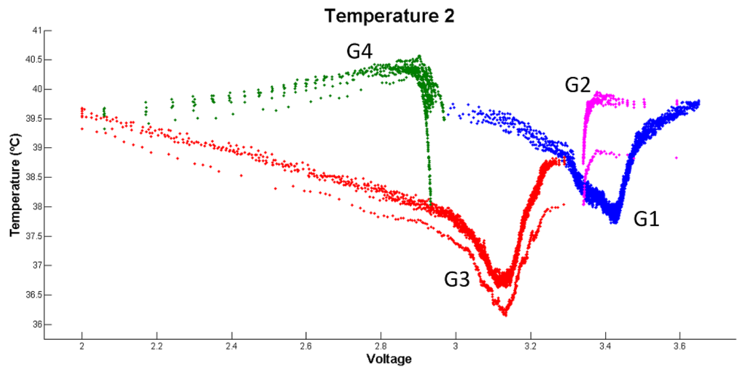

Fig. 6. Temperature 2 vs. Battery voltage

4 Results

The results of the clustering algorithm was compared with the real state asigned during the test. Due to the fact that the dataset has the correct properties to use K-Means, the clustering achieved was 100% of correct assignation. This fact allows the model approach not to need to know the cycle state for the data.

14 different ANN-MLP were tested for each cluster, with a number of neurons in the hidden layer from 2 to 15. In all cases, these neurons have a Tan-sigmoidal transfer function, and the output layer neuron has a linear transfer function.

10 different polynomial regressions were tested for each cluster. For this technique, the degrees of the polynomial used were from degree 1 to degree 10.

The ensemble learning method used was 'LSBoost', it was 5000 trained cycles and a regression tree algorithm; one ensemble was training for each cluster.

All the models were compared by using the Mean Square Error (MSE) as the efficiency measurement. The testing data are only used to calculate the MSE, not for training any model.

In table 2 the lowest MSE achieved appears for each algorithm. Table 3 shows the best regression technique and its configuration for each cluster. Even so, the selection takes the computational cost into account when the MSE for different techniques are close.

It is remarkable that the best MSE achieved without clustering the dataset was, at least, twice worse than the worst result reached with the proposal. In the capacity model, the average MSE with clustering is 0.0726, and without it is over than 236.

5 Conclusions

Very good results have been obtained in general terms with the novel approach proposed in this research. The average of the MSE is 0.0590 varying form 0.0014 and 0.2541 for the different variables depending of the cycle state. It is possible

Variable	Cycle state	ANN-MLP MSE	Polynomial MSE	Ensemble MSE
Temperature 1	Charge (1)	0.0025	0.0413	0.0028
Temperature 1	Rest (2)	0.0862	0.0884	0.1023
Temperature 1	Discharge (3)	0.0344	0.0642	0.0361
Temperature 1	Rest (4)	0.0881	0.1270	0.1024
Temperature 2	Charge (1)	0.0056	0.0413	0.0059
Temperature 2	Rest (2)	0.0761	0.0782	0.0895
Temperature 2	Discharge (3)	0.0316	0.0577	0.0335
Temperature 2	Rest (4)	0.0940	0.1324	0.1098
Capacity	Charge (1)	0.0329	13.7110	0.0516
Capacity	Rest (2)	0.0016	0.0014	0.0016
Capacity	Discharge (3)	0.2541	2.7005	0.2933
Capacity	Rest (4)	0.0021	0.0019	0.0023

Table 2. Best MSE for each regression algorithm

Variable	Cycle state	Model	MSE
Temperature 1	Charge (1)	ANN-MLP, 5 neurons	0.0025
Temperature 1	Rest (2)	ANN-MLP, 2 neurons	0.0862
Temperature 1	Discharge (3)	ANN-MLP, 5 neurons	0.0344
Temperature 1	Rest (4)	ANN-MLP, 3 neurons	0.0881
Temperature 2	Charge (1)	ANN-MLP, 8 neurons	0.0056
Temperature 2	Rest (2)	ANN-MLP, 2 neurons	0.0761
Temperature 2	Discharge (3)	ANN-MLP, 5 neurons	0.0316
Temperature 2	Rest (4)	ANN-MLP, 3 neurons	0.0940
Capacity	Charge (1)	ANN-MLP, 12 neurons	0.0329
Capacity	Rest (2)	Polynomial 1	0.0014
Capacity	Discharge (3)	ANN-MLP, 7 neurons	0.2541
Capacity	Rest (4)	Polynomial 1	0.0019

Table 3. MSE for the best methods

to predict the value of the SOC in real time for the capacity confirmation of the battery test. This model could be used to ensure a good power cell test, for example, by detecting when a test provides wrong results.

The results achieved with the hybrid model increase the whole efficiency of the approach because each model was trained only for a group of the dataset. For the regression, the best approximation has been obtained with MLP in all cases, with the exception of SOC prediction for Rest 2 and 4. For this two cases, the best MSE is reached with Polynomial Regression.

In other terms, the cell temperature is a critical parameter that is so significant of the device health. The temperatures in the battery were included into the model to predict a deviation from the normal settings in the test. Thus, it is possible to detect deviations in this sense, too.

References

1. Chukwuka, C., Folly, K.: Batteries and super-capacitors. Power Engineering Society Conference and Exposition in Africa (PowerAfrica), 2012 IEEE (July 2012) 1–6
2. Qian, H., Zhang, J., Lai, J.S.: A grid-tie battery energy storage system. Control and Modeling for Power Electronics (COMPEL), 2010 IEEE 12th Workshop on (June 2010) 1–5
3. Chaturvedi, N., Klein, R., Christensen, J., Ahmed, J., Kojic, A.: Modeling, estimation, and control challenges for lithium-ion batteries. American Control Conference (ACC) (June 2010) 1997–2002
4. Vukosavic, S.: Electrical Machines. Power Electronics and Power Systems. Springer (2012)
5. Konig, O., Jakubek, S., Prochart, G.: Battery impedance emulation for hybrid and electric powertrain testing. Vehicle Power and Propulsion Conference (VPPC), 2012 IEEE (Oct 2012) 627–632
6. Mark, J., Goldberg, M.A.: Multiple regression analysis and mass assessment: A review of the issues. Appraisal Journal **56**(1) (1988) 89
7. Jin, R., Cho, K., Hyun, C., Son, M.: Mra-based revised CBR model for cost prediction in the early stage of construction projects. Expert Systems with Applications **39**(5) (2012) 5214–5222
8. Ho, L.H., Feng, S.Y., Lee, Y.C., Yen, T.M.: Using modified IPA to evaluate suppliers performance: Multiple regression analysis and DEMATEL approach. Expert Systems with Applications **39**(8) (2012) 7102–7109
9. Cho, Y., Awbi, H.B.: A study of the effect of heat source location in a ventilated room using multiple regression analysis. Building and Environment **42**(5) (2007) 2072–2082
10. Nieves-Acedo, J., Santos-Grueiro, I., Garcia-Bringas, P.: Enhancing the prediction stage of a model predictive control systems through meta-classifiers. Dyna **88**(3) (2013) 290–298
11. Jang, H., Topal, E.: Optimizing overbreak prediction based on geological parameters comparing multiple regression analysis and artificial neural network. Tunnelling and Underground Space Technology **38**(0) (2013) 161–169
12. Alvarez-Huerta, A., Gonzalez-Miguelez, R., Garcia-Metola, D., Noriega-Gonzalez, A.: Failure detection and prediction in wind turbines by using scada data. Dyna **86**(4) (2011) 467–473
13. Martínez-Rego, D., Fontenla-Romero, O., Alonso-Betanzos, A.: Efficiency of local models ensembles for time series prediction. Expert Syst. Appl. **38**(6) (June 2011) 6884–6894
14. Cherif, A., Cardot, H., Boné, R.: SOM time series clustering and prediction with recurrent neural networks. Neurocomput. **74**(11) (May 2011) 1936–1944
15. Ghaseminezhad, M.H., Karami, A.: A novel self-organizing map (SOM) neural network for discrete groups of data clustering. Appl. Soft Comput. **11**(4) (June 2011) 3771–3778
16. Crespo-Ramos, M.J., Machón-González, I., López-García, H., Calvo-Rolle, J.L.: Detection of locally relevant variables using SOM-NG algorithm. Engineering Applications of Artificial Intelligence **26**(8) (2013) 1992–2000
17. Calvo-Rolle, J.L., Casteleiro-Roca, J.L., Quintián, H., Meizoso-Lopez, M.C.: A hybrid intelligent system for PID controller using in a steel rolling process. Expert Systems with Applications **40**(13) (2013) 5188–5196

18. Wikipedia: Lithium-ion battery - wikipedia, the free encyclopedia (2014) [Online; accessed 30-January-2014].
19. Sparacino, A., Reed, G., Kerestes, R., Grainger, B., Smith, Z.: Survey of battery energy storage systems and modeling techniques. Power and Energy Society General Meeting, 2012 IEEE (July 2012) 1–8
20. : PNGV Battery Test Manual. (2 2001)
21. : LiFeBATT X-1P 8Ah 38123 Cell. (3 2011)
22. Qin, A., Suganthan, P.: Enhanced neural gas network for prototype-based clustering. Pattern Recogn. **38**(8) (August 2005) 1275–1288
23. Ye, J., Xiong, T.: SVM versus least squares SVM. Journal of Machine Learning Research - Proceedings Track **2** (2007) 644–651
24. Kaski, S., Sinkkonen, J., Klami, A.: Discriminative clustering. Neurocomputing **69**(1-3) (2005) 18–41
25. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006)
26. UYSAL, I., GÖVENIR, H.A.: An overview of regression techniques for knowledge discovery. The Knowledge Engineering Review **14** (12 1999) 319–340
27. Mozaffari, A., Azad, N.L.: Optimally pruned extreme learning machine with ensemble of regularization techniques and negative correlation penalty applied to automotive engine coldstart hydrocarbon emission identification. Neurocomputing (0) (2013) –