



## Data driven fault detection and isolation: a wind turbine scenario

### Detección y aislamiento de fallas guiadas por datos: escenario turbina de viento

Rubén Francisco Manrique Piramanrique\*, Jorge Sofrony Esmeral\*

**Fecha de recepción:** 30 de noviembre de 2013

**Fecha de aceptación:** 5 de diciembre de 2014

**Citation / Para citar este artículo:** Manrique Piramanrique, R. F., & Sofrony Esmeral, J. (2015). Data driven fault detection and isolation: a wind turbine scenario. *Revista Tecnura*, 19(44), 71-82. doi:<http://dx.doi.org/10.14483/udistrital.jour.tecnura.2015.2.a05>

#### Abstract

One of the greatest drawbacks in wind energy generation is the high maintenance cost associated to mechanical faults. This problem becomes more evident in utility scale wind turbines, where the increased size and nominal capacity comes with additional problems associated with structural vibrations and aeroelastic effects in the blades. Due to the increased operation capability, it is imperative to detect system degradation and faults in an efficient manner, maintaining system integrity, reliability and reducing operation costs. This paper presents a comprehensive comparison of four different Fault Detection and Isolation (FDI) filters based on "Data Driven" (DD) techniques. In order to enhance FDI performance, a multi-level strategy is used where: (i) the first level detects the occurrence of any given fault (detection), while (ii) the second identifies the source of the fault (isolation). Four different DD classification techniques (namely Support Vector Machines, Artificial Neural Networks, K Nearest Neighbors and Gaussian Mixture Models) were studied and compared for each of the proposed classification levels. The best strategy at each level

could be selected to build the final data driven FDI system. The performance of the proposed scheme is evaluated on a benchmark model of a commercial wind turbine.

**Keywords:** data mining, fault detection, wind energy.

#### Resumen

Uno de los mayores inconvenientes presentes en la generación de energía eólica son los altos costos de mantenimiento asociados a fallas mecánicas. Este problema se hace más evidente en las turbinas de viento de escala industrial, en donde incrementos en el tamaño y la capacidad nominal traen consigo problemas adicionales asociados a vibraciones estructurales y efectos aeroelásticos en las hojas. Debido al incremento en la capacidad de operación, es imprescindible detectar de manera eficiente fallas y degradaciones en el sistema, garantizando la integridad, su fiabilidad y reduciendo los costos de operación. Este artículo presenta un sistema para la detección y aislamiento de fallas (FDI), basado en técnicas "Guiadas por los datos" (*Data driven*, abreviado DD). La arquitectura propuesta es una estrategia de varios niveles en donde: (i) el primer nivel

\* Electrical Engineer, M.Sc. in Computer Science, Assistant Professor at Universidad Sergio Arboleda. Bogotá, Colombia. Contacto: [ruben.manrique@usa.edu.co](mailto:ruben.manrique@usa.edu.co)

\* Electrical Engineer, M.Sc. and Ph.D. in Control Systems, Associate Professor at Universidad Nacional de Colombia, Bogotá, Colombia. Contacto: [jsofronye@unal.edu.co](mailto:jsofronye@unal.edu.co)

detecta la ocurrencia de una falla (detección), mientras que (ii) el segundo identifica su origen (aislamiento). Se estudiaron y compararon cuatro técnicas de clasificación para cada uno de los niveles (Máquinas de Vectores de Soporte, Redes Neuronales Artificiales, K vecinos cercanos y Mezcla de Gaussianas). La

mejor estrategia en cada nivel fue seleccionada para construir el sistema FDI. El rendimiento del sistema propuesto se evalúa en un modelo de referencia de una turbina eólica de escala comercial.

**Palabras clave:** energía eólica, detección de fallas, minería de datos.

## INTRODUCTION

Energy generation is one of humanities greatest concerns, as most social activity has been built around its consumption. This highlights the fact that we need to introduce new sources of “green” energy that are reliable, accessible, and economically attractive. Within this context, wind energy has become one of the most popular options, observing a market growth of approximately 24% in the last decade (GWEC Global Energy Council, 2011). Wind power and its transformation into electrical energy is a simple idea that comes along with many challenges. The underlying idea is that wind power absorption is proportional to the square of the blade length and to the wind velocity cubed. As the demand of installed capacity increases, so does the need to design generators with higher rated capacity. This may only be achieved either by increasing the wind speed (which is generally unfeasible) or increasing the size of the blades. At these sizes a significant increases in the structural loads and vibrations are introduced. If these increase in load is not properly addressed the system may be more prone to malfunction in its mechanical components, and consequently less cost efficient. Although the efficiency of wind turbines has improved greatly, their reliability has decreased. This leads to an increase in maintenance costs, which translates into longer down times and captured energy losses.

Modern active control techniques can optimize efficiency and increase system reliability. Nonetheless, faults are likely to occur and have a negative impact on system performance, or even

result in catastrophic events if not detected on time. In addition, wind turbines are located in remote locations, under extreme environmental conditions, making it difficult to implement preventive maintenance plans. It has been reported (Walford, 2006) that approximately 10%-15% of the total revenues due to wind energy generation has to be reinvested into maintenance. Analytical Fault Detection and Isolation (FDI) may offer a cost efficient solution that allows for early fault detection. It is believed that FDI techniques, and the early detection of faults, can reduce the mean down time of wind turbines and repair costs associated with their operation. Industrial demand for such FDI systems has seen an exponential growth in recent years, and has been accompanied by higher performance and reliability demands. FDI techniques have become an essential component in the development of intelligent autonomous systems where it is necessary to constantly monitor the health state (detect, isolate and estimate severity of the failure mode) of the system.

Model based techniques are the most common approach to building analytic FDI systems, where many techniques have been proposed including Kalman filters (Bergantino & S, 2009), unknown input observers (Chen & Saif, 2006), and  $H_\infty/H_2$  filters (Verhaegen, 2008), have been used. The underlying idea behind model based FDI is to use a mathematical model of the system as the source of redundant information, and produce estimates of the systems measured outputs, this model is usually developed based on some fundamental understanding of the physics of the process. The disadvantage of this approach is the need for an

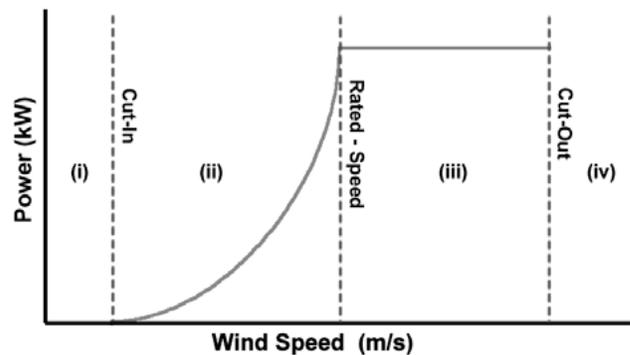
accurate model of the process; in other words, model based techniques are affected by model uncertainties, plant disturbances and measurement noise. An alternative approach known as “Data Driven” (DD) FDI uses a collection of measured data to unearth patterns of normal and faulty behavior. DD FDI is based on the advances in areas such as data mining, machine learning and computational intelligence. Some of the most popular DD FDI techniques used Artificial Neural Networks (Zhou, 2004), (Kusiak & Shah, 2006) and Bayesian Networks (Sylvain & Kobi, 2010), amongst others.

This article presents the application of data mining techniques to a Fault Detection and Isolation problem setup. The main contribution is the proposition of a multi-level architecture where the first level must evaluate the existence of a fault, while the objective of the second level is to isolate the source of the fault. This strategy, in combination with a FCTW, proved to enhance performance of the FDI filters, hence making them comply with industrial standards and readily available for the practitioner. The paper starts with a brief outline of the classification techniques used; it follows by describing the DD FDI architecture proposed and its application to a benchmark model of a utility scale wind turbine as proposed in (Odgaard, 2009); the DD FDI system is constructed and simulation results are presented to highlight the usefulness of the data driven approach; finally some concluding remarks are given.

## WIND TURBINE CONFIGURATION

The basic wind turbine configuration is as follows. The nacelle contains main components including the generator and gearbox, and sits on top of the tower structure. The blades are connected to the rotor, which in turn is connected to the generator via the gearbox. The nacelle contains a yaw motor that allows the blades to face the wind. Large utility scale generators are generally variable speed and have mechanisms that change the blade pitch angle

and control the lift produced by the wind. Wind turbines have additional actuators and sensors that can be used for control. The generator torque load can be used to dictate the amount of electrical power absorbed from the mechanical system, and act as a braking system to control the acceleration of the rotor; a braking system can set the rotor to a full stop. The main measurements are the rotor and generator speeds and blade pitch angles. As one of the main premises is that the system must remain low cost, high cost and redundant sensors are generally avoided, which entail additional challenges to the closed loop control and FDI problems.



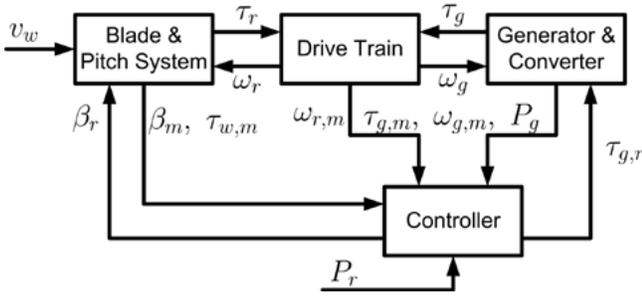
**Figure 1.** Ideal wind turbine power curve and operating zones.

**Source:** own work.

The turbine operates at four distinct regions, figure 1: (i) cut in speed, where winds are insufficient; (ii) the interim region, i.e. the region between the cut in and rated wind speed; (iii) maximum power capture region, which starts at the minimum rated wind speed; (iv) cut out speed, where wind speeds are too high. The general control objective is to maximize power absorption while operating in region (ii) and minimize structural loads when operating in region (iii); regions (i) and (iv) are not contemplated.

The data set was generated from simulation results of a “benchmark” wind turbine model built

for the development and analysis of FDI systems (Odgaard, 2009). In figure 2 the components of the model used and the relationship between them are represented. This model is a simplification of a three blade, utility scale (horizontal) wind turbine with a rated power of 4.8 MW. Each blade has independent pitch motion, hence the control system provides three separate pitch command signals, i.e  $\beta_{i,r}$  where  $i = [1,2,3]$ ; the actual pitch position is measured by sensors that provide signals  $\beta_{i,m}$ . Although the model has (dual) physical redundancy for the pitch sensors, the aim of this paper is to analyze the performance of an analytical DD FDI system so we will eliminate redundant measurements.



**Figure 2.** Wind Turbine Model: System Interconnection.

**Source:** (Odgaard, 2009).

The torque transmitted to the generator and the rotor is denoted  $\tau_g$  and  $\tau_r$  respectively, while the

angular velocity of the generator and the rotor are denoted by  $\omega_g$  (rad/s) and  $\omega_r$  (rad/s) (respectively). The complete set of measurements (in addition to pitch angle measurements) includes the transmitted torques, angular velocities and the wind speed ( $v_w$ ). The captured power is approximately given by equation (1).

$$P_r = \tau_{aero} \omega_r = \frac{1}{2} \rho A v_w^3 C_p(\lambda, \beta) \quad (1)$$

where  $\tau_{aero}$  (N) is the aerodynamic torque,  $\omega_r$  is the rotor speed,  $\rho$  (kg/m<sup>3</sup>) is the air density,  $A$  (m<sup>2</sup>) is the area swept by the rotor,  $v_w$  (m/s) is the wind speed.  $C_p$  is the power coefficient which represents how much power available in wind is captured.  $C_p$  is a function of blade pitch angle  $\beta$  (deg) and tip speed ratio  $\lambda = \frac{\omega_r R}{v_w}$ ,  $R$  (m) is the rotor radius. The pitch actuators on the system are represented by a transfer function  $G_{act}(s)$  (equation (2) with  $\zeta = 0.6$  and  $\omega_n = 11.11$ ).

$$\frac{\beta(s)}{\beta_{ref}(s)} = G_{act}(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (2)$$

The generator's dynamics are modeled with a first-order transfer function, equation (3).

$$\frac{\tau_g(s)}{\tau_{g,ref}(s)} = G_{gen}(s) = \frac{\alpha_g}{s + \alpha_g} \quad (3)$$

The drive train and generator are modeled by a two mass model, equation (4).

$$\begin{bmatrix} \dot{\omega}_r \\ \dot{\omega}_g \\ \dot{\theta}_\Delta \end{bmatrix} = \begin{bmatrix} -(B_{dt} + B_r) & B_{dt} & -k_{dt} \\ J_r & N_g J_r & J_r \\ \eta_{at} B_{dt} & -\eta_{at} B_{dt} - B_g N_g^2 & \eta_{at} K_{dt} \\ N_g J_g & N_g^2 J_g & N_g J_g \\ 1 & -1/N_g & 0 \end{bmatrix} \begin{bmatrix} \omega_r \\ \omega_g \\ \theta_\Delta \end{bmatrix} + \begin{bmatrix} -1/J_r & 0 \\ 0 & 1/J_g \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tau_{aero} \\ \tau_g \end{bmatrix} \quad (4)$$

Where  $\theta_{\Delta}$  is the drivetrain torsion,  $J_r$  and  $J_g$  are the rotor and generator inertia,  $B_r$  and  $B_g$  are the viscous damping of the rotor and generator.  $B_{dt}$  and  $K_{dt}$  are the damping and stiffness coefficients of the drivetrain flexibility,  $N_{g_i}$  is the gearbox ratio and  $\eta_{dt}$  is the efficiency of the drive train. If the turbine is operating in Region 3, a discrete PI controller generates the pitch commands references to maintain the rated rotor speed. The generator torque is set as  $\tau_{g,ref} = P_{rated}/\omega_g$ .

## APPLICATION OF DATA MINING TECHNIQUES

The fault diagnosis and isolation problem involves determining the failed component (isolation) once a fault has been detected (detection). The ability to isolate faults can be made “simple” at the expense of additional, physically redundant, sensors and increased costs, thus it is of interest to develop techniques that use analytical (redundant) information. Although the most common approach to analytical FDI is to use model-based techniques, the construction of a high fidelity model is a cumbersome task; data mining techniques allow the detection and isolation of system faults by only using input/output data.

From a data mining perspective, classification is the task of finding (training) an objective function  $f(x)$  which assigns one of the predefined class labels  $y$ , to each set of the attributes  $x$ ; the training data is the collection of records (examples), characterized by a tuple  $(x,y)$ . For the case of an FDI system,  $y$  may take values associated with each of the fault modes and  $x$  would be the vector of measured signals of the system. Although there is a broad range of classification strategies, this paper will focus on four of them: Support Vector Machines (SVM), Artificial Neural Networks (ANN), Bayesian Classifiers (i.e Naive Bayes with Gaussian Mixture Models) and K Nearest Neighbors (KNN). A brief description of each of these techniques will be provided next.

*Support vector machines:* Vladimir Vapnik first introduced SVM within the context of binary

classification problems. The basic idea behind SVM is to find an optimal hyperplane that maximizes the separation margin between the classes. Finding this hyperplane is equivalent to solving a constrained optimization problem whose solution is a linear combination of training examples that lie on the edge of the margin; these are known as “support vectors”. The mathematical description of the algorithm is extensive and can be found in (Scholkopf & Smola, 2001).

*Artificial neural networks:* These are inspired by a simplification of the biological neural system, which consists of simple computing elements (called neurons) connected through synaptic weights. The multilayer perceptron (MLP) is the most applied architecture and consists of an input layer, at least one hidden layer, and an output layer. The input to a given neuron is a linear combination of the output of neurons from the previous layer, where each input is weighted by their respective synaptic weight. Associated to each neuron, there exists a nonlinear activation function, which transforms the linear combination of inputs into the neurons output, which in turn is an input to the neurons in the next layer. For the application proposed, the number of hidden layers and neurons was determined experimentally, where sigmoidal activation functions achieved the best results. The ANN is trained using a back propagation with gradient descent algorithm as detailed in (Bishop C. M., 2006).

*K nearest neighbors:* The K Nearest Neighbors (KNN) approach consists of a supervised classifier based on neighborhoods. The underlying idea is that a new sample belongs to the class to which the largest numbers of (k nearest) neighbors belong. One of the challenges is to choose appropriately the number (i.e.  $k$ ) of neighbors to be analyzed; if  $k$  is too large it may be difficult to discriminate from global tendencies, but if  $k$  is small classification may be imprecise because of the lack of data. Note that KNN depends greatly on the way distance is calculated, thus it is possible to have different classifiers by using different distance metrics. The

classifier proposed in this paper uses the Euclidean norm, where the Euclidean norm of a vector  $x$  is given by  $\|x\| = \sqrt{x^T x}$ . The distance between two vectors,  $x_i$  and  $x_j$ , is defined as  $dist(x_i, x_j) = \|x_i - x_j\|$  as the metric for distance.

**Bayesian classifiers:** Bayesian classifiers are based on probability theory, where classification is made depending on which class has the greatest probability of occurrence or the lowest risk (Duda, Hart, & Stork, 2001). Assume that the problem considered is the classification of a set of attributes as one of  $K$  different classes. The vector of attributes (features) is denoted as  $\mathbf{x} = [x_1, x_2, \dots, x_d]^T$  where  $d$  is the dimension of the vector and the probability that the vector  $x$  belongs to the class  $y_k$  is  $p(y_k|\mathbf{x})$  (known as *posterior* probability). The attribute set  $x$  belongs to the class with the highest posterior probability, where can be calculated according to Bayes' formula, equation (5).

$$p(y_k|\mathbf{x}) = \frac{p(\mathbf{x}|y_k)p(y_k)}{p(\mathbf{x})} \quad (5)$$

Where  $p(\mathbf{x}|y_k)$  is the probability density function of class  $y_k$  and  $p(y_k)$  is the *prior* probability of the class. If the *prior* probabilities are not currently known, they can be estimated from the proportions of the class in the training set (Tan, Steinbach, & Kumar, 2005). A feature vector  $\mathbf{x}$ , belongs to the class  $y_k$  if  $p(y_k|\mathbf{x})$  is has highest posterior probability of all the  $K$  classes:  $p(y_k|\mathbf{x}) > p(y_j|\mathbf{x}) \quad \forall j \neq k$ . The biggest challenge when implementing Bayesian classifiers is to estimate the probability density function (PDF)  $p(\mathbf{x}|y_k)$ . In practice, this PDF is always unknown and must be calculated from the available information, i.e. the '*training data*'. Two approaches, namely Naive Bayes and the Mixture of Gaussians, will be presented next.

The Naive Bayes (NB) approach estimates  $p(\mathbf{x}|y_k)$  assuming that the attributes are conditionally independent for a given class  $k$ . To estimate the conditional probabilities  $p(x_i|y_k)$ , we can assume a certain form of probability distribution and calculate the parameters of the distribution using the training data (Tan, Steinbach, & Kumar, 2005).

In many cases an exact density function can be difficult to obtain or not follow any particular distribution function. If this is the case, it is possible to use a linear combination of Gaussian distributions to represent (i.e. model) the "real" PDF of each class (Gaussian Mixture Models). Each distribution function is the modeled via equation (6).

$$N(\mathbf{x}, \mu, \Sigma) = \frac{1}{2\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} e^{-1/2(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)} \quad (6)$$

Where  $\mathbf{x}$  is the vector of features,  $\mu$  is the mean vector,  $\Sigma$  is the covariance matrix and  $|\Sigma|$  is its determinant. In our particular case, the classifier was constructed under the following premises: *i*) the probabilities of occurrence of each class, i.e.  $p(y_k)$  for  $k = [1, \dots, K]$ , are known and equal to  $1/K$ ; *ii*) the probability density functions  $p(\mathbf{x}|y_k)$  for  $k = [1, \dots, K]$ , are constructed using a linear combination of  $n$  multivariate Gaussian distributions; *iii*) since there is no optimal way to choose  $n$  (Sylvain & Kobi, 2010), its value was chosen experimentally; *iv*) the GMM parameters are calculated using an Expectation Maximization Algorithm (Alpaydin, 2004).

## METHODOLOGY

The data set corresponds to a simulation of 3700 seconds with the model specified above and sampling time  $T_s = 0,01$  seg. A total of seven fault modes were studied, originally proposed by (Odgaard, 2009), each of which was triggered at different time intervals. *Fault 1*: Fixed value in pitch 1 position sensor measurement, *Fault 2*: Scaling error in pitch 2 position sensor measurement, *Fault 3*: Fixed value in Pitch 3 position sensor measurement, *Fault 4*: Fixed value in rotor speed sensor measurement, *Fault 5*: Scaling error in rotor and generator speed sensor measurements, *Fault 6*: Degraded hydraulic pitch system response pitch actuator 2 and *Fault 7*: Air in Oil Failure pitch actuator 3.

The signals of the monitoring system are wind speed, pitch position, generator speed, rotor speed and generated power (as shown in

table 1) and constitute our initial set of attributes. The classifiers were trained according to the algorithms discussed earlier in the paper, and results are analyzed in terms of detection times

and false alarm clear out time. To evaluate the models constructed from the perspective of data mining (i.e. prediction levels), the following metrics are defined (equations (7) and (8)).

$$\text{Accuracy} = \frac{\text{correctly classified records (normal + fault)}}{\text{Total number of records}} \times 100\% \quad (7)$$

$$\text{Sensitivity} = \frac{\text{Fault records correctly classified}}{\text{Total number of fault records}} \times 100\% \quad (8)$$

**Table 1.** Attribute Set.

Complete Attribute Set			
Variable	Rate	Measurement	Description
$v_s$	$\Delta(v_s)$	WindSp [m/s]	Wind speed sensor
$\omega_r$	$\Delta(\omega_r)$	RotorSp [rad/s]	Rotor speed measurement 1
$\omega_g$	$\Delta(\omega_g)$	GenSp [rad/s]	Generator speed measurement 1
$\tau_r$	$\Delta(\tau_r)$	GenTorq [kN m]	Measured Generator torque
$\tau_g$	$\Delta(\tau_g)$	GenPower [kNm/s]	Measured Generator Power
$B_{1m}$	$\Delta(B_{1m})$	Bl1Pitch [deg]	Pitch angle 1 measurement 1
$B_{2m}$	$\Delta(B_{2m})$	Bl2Pitch [deg]	Pitch angle 2 measurement 1
$B_{3m}$	$\Delta(B_{3m})$	Bl3Pitch [deg]	Pitch angle 3 measurement 1

Source: own work.

### Data-Driven fault detection and isolation system training

DD FDI systems use measured signals as attributes to build the training set, thus the quality of the classifiers depends greatly on the quality of the training set used. In order to increase the classification model's accuracy, three actions were performed on the data: data sampling, outlier elimination and attribute (feature) creation.

*Data sampling:* Sampling is the process of selecting a representative (reduced) set that still allows a thorough examination of the entire data. Sampling is needed in order to allow the abstraction of a complex problem, as well as to acquire a sub set that infers information from a larger data set. It is widely accepted that a fairly

modest sized sample can sufficiently characterize a much larger population. The usefulness of the sample is determined by two characteristics: the size and the "quality" of the sample. The sample size should not be too small since it may misrepresent the entire data set, or too large that it overloads the computational algorithms.

In order to generate a suitable training set, it was necessary to reduce the total number of data registers and balance the proportion of "normal" and "faulty" operation events. There exist several procedures that aid in the process of data set sampling. In this paper we followed a simple procedure (random sampling) that allowed a *balanced reduction* of the class labeled as "normal". A (simple) random sample of size

$n$  consists of  $n$  individuals from the population chosen in such a way that every individual has an equal chance of being selected. The total number of data registers used for training was 390,000, where 75% correspond to data representing the normal operation of the system, while the remaining registers correspond to faulty behavior. Through random sampling, this number was reduced to 197,000 registers and a balanced data set was obtained.

*Outlier elimination:* An outlier is generally considered to be a data point that is far outside the normal behavior of a variable or population. Outliers can have deleterious effects on statistical analyses since they generally serve to increase error variance and reduce the relevance of statistical tests. Additionally they can seriously bias or influence estimates that may be of substantive interest. It is therefore desirable to eliminate data that may deteriorate detection performance, i.e. outlier elimination.

Several formal statistical tests have been devised to determine if a certain value can be considered as an outlier. In its simplest form, box plots present five measurements (the minimum, the lower quartile, the median, the upper quartile and the maximum), all in a visual display. The lower quartile has 25% of the sample values below and 75% above it. The upper quartile has 25% of the sample values above it and 75% below. The middle quartile is the median and the middle half of the sample lies between the upper and lower quartile. The distance between the upper and lower quartile is called the interquartile range. Any observation outside the interquartile range is considered a potential outlier. After analyzing data box plots, it was possible to identify and eliminate a set of outliers associated with the turbine's start up process and can therefore be ignored. The outlying data provides no useful information and may correspond to an event that we do not wish to classify.

*Creating new attributes:* From a viewpoint of data mining, it is possible to increase sensitivity

to a particular class and accuracy of a classifier by creating new attributes (Tan, Steinbach, & Kumar, 2005). For example, fixed value sensor faults may be difficult to detect if the faulty measurement is close to steady state, normal operation value. This problem can be solved by creating a new set of attributes that takes into account the "velocity" of the measurements, and not only its value. Although it may be possible to find more robust attributes, we want to avoid the pre-processing of input data as much as possible; hence the rate of change is a suitable choice. By defining  $\Delta(x_i) = |x_i - x_{i-1}|$  as the rate of change of the measurement  $x$  a new set of attributes is constructed.

As an exploratory exercise, and with the objective of determining the quality of the FDI models that could be obtained using the previously mentioned techniques, four classifiers were constructed. The training set consists of a family of attributes (see table 1) and a flag that determines the occurrence of a given fault; the flag will take a zero value when no fault occurs (i.e. normal behavior), and will take a value within the set  $\{1,2,3,4,5,6,7\}$  in the event of fault occurrence (i.e. the flag values is 1 where Fault 1 is present, and so on). At this stage a full FDI system is constructed for each one of the methods proposed.

To evaluate the performance of each FDI system, a total 50 simulations were performed per technique proposed. Each simulation had duration of 3700 seconds, exhibiting all of the faults at different times. It is important to mention that simulations sequences were different from the training set, and that no preprocessing was performed on the data. The results obtained are presented in table 2.

Observing the general results presented in table 2, it is possible to draw the following conclusions. First, it is important to highlight the need to generate suitable training sets by using outlier elimination techniques and appropriate data sampling (an average 10% degradation was observed when no preprocessing was performed). This will reduce the risk of having the model memorize the training set

**Table 2.** Average Accuracy and Sensitivity.

Algorithm	Sensitivity (%)							Accuracy (%)
	F1	F2	F3	F4	F5	F6	F7	
KNN	93.20	70.12	85.23	68.21	95.12	41.12	57.12	65.42
NB	91.70	85.80	84.20	82.60	93.20	60.90	59.80	80.72
ANN	82.60	84.00	82.70	76.10	87.80	67.20	62.50	78.63
SVM	86.70	88.80	80.20	85.60	98.20	68.90	67.80	74.89

Source: own work.

and avoid any unwanted bias caused by abnormal operation states (i.e. turbine startup). Second, Faults 1 through 5 are abrupt in nature and their detection presents acceptable levels of accuracy and sensitivity, and a low number of false alarms. Unfortunately, for Faults 6 and 7, low sensibility is exhibited and a high number of false alarms are present. Because of this the presented strategy is not suitable for the FDI application at hand and it is necessary to investigate new architectures. Third, it was observed that incipient faults were more difficult to detect due to their slow growing nature. In fact, it was observed that ANN and SVM techniques provided better detection properties for incipient faults, thus a reliable DD FDI system may require the application of more than one data mining technique.

In general, probabilistic methods tend to present better classification performance, and this is supported by the fact that the Naive Bayes method presented the best accuracy and sensitivity indicators of the four algorithms tested. Nonetheless, in order to obtain good results it is necessary to have an accurate (or good enough) PDF. Initially, the task of finding a PDF for detection (determining state of Fault/No Fault) may seem to be relatively easy because the problem at hand is that of binary classification. As the requirements scale up to identification, finding an accurate PDF may become rather complex, thus the need for more advance techniques such as Mixed Gaussian Models to determine an approximate, but accurate enough PDF.

Data driven techniques have proved to be a promising approach to fault detection and isolation, but the previous result also stressed the fact that more heuristic architectures are needed in order to reach the performance levels required for industrial applications. Having this in mind, the next section proposes a *Multi Level* architecture that deals with the problem of detection and the problem of isolation in a sequential manner.

## MULTI LEVEL DD FDI

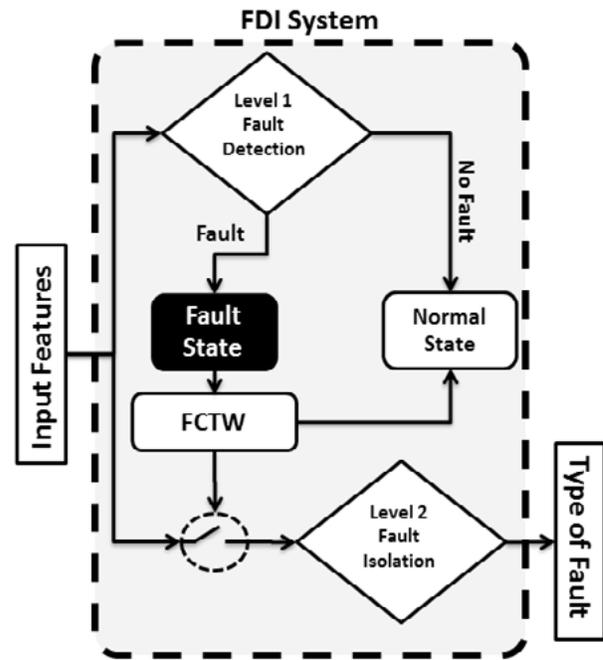
The FDI system proposed uses a *two level* detection architecture: the first classification level is responsible for detecting the fault and distinguishing faulty from normal behavior; the second level is responsible of isolating the source of the failure. The FDI architecture is summarized in figure 3. Details on the construction of each classification level are discussed next.

*Level 1:* This level is posed as a binary problem where the main classes are "Fault" and "Normal". It is important to note that at this stage it is imperative to recognize states of failure, hence the classifier is trained to have high sensitivity at the expense of partial degradation of the general accuracy. As a consequence, this level has a high number of false alarms, so in order to enhance accuracy a Fault Counter on Time Windows (FCTW), as proposed in (Lipnickas A. , 2006), was implemented. The main objective is to verify the "persistence" of the faulty behavior, so the classifier must identify the "Fault" persistently

over a given period of time. The size of the time window was determined experimentally by considering that there is a trade off between accuracy and detection time. It should be noted that this approach is not suitable for intermittent faults with long periods of occurrence.

The average performance over 50 test runs is summarized in table 3. The accuracy is, in general, higher than 80%, with sensitivity above 90% for SVM, ANN and GMM. At this level we are interested in a classification model with high sensitivity, which usually presents a higher number of false alarms. It can be note no model has an accuracy greater than 90%, indicating that there is a low number of undetected faults, but a considerable number of false alarms. The table also shows the percentage of false alarms (FA %) and the recovery time after its occurrence (FA Clear Time).

*Level 2:* This stage consists of a multi class classification problem, where all the possible occurrences of fault are a given class. At this level the training set is constructed with data in a state of "Fault", and the main goal is to isolate the



**Figure 3.** Multilayer Classification Architecture for FDI.

Source: own work.

**Table 3.** Accuracy and Sensitivity Level 1 using FCTW.

Algorithm	Accuracy (%)	Sensitivity (%)	FA (%)	FA Clear Time
SVM+FCTW	97.290	89.32	1.010	3,5Ts
ANN+FCTW	97.10	88.14	1.012	3,5Ts
KNN+FCTW	91.18	79.97	5.63	4,3Ts
GMM+FCTW	92.27	84.74	5.30	4,1Ts

Source: own work.

**Table 4.** Accuracy and Sensitivity Level 2.

Algorithm	Accuracy (%)	Sensitivity (%)						
		F1	F2	F3	F4	F5	F6	F7
SVM	84.2	96	80.5	96.1	87.1	82.8	60.2	72.2
ANN	89.5	94.1	85.2	99	86	79.1	65.9	78.7
KNN	59.2	84.2	65.2	91.2	79.2	45.7	38,3	39.4
GMM	90.2	95.3	75.9	93.2	86.3	88.6	71.2	88.5

Source: own work.

source of the fault. All the implemented strategies are multi class classifiers, i.e. there are a total of 7 classes that need to be identified. The results of the validation tests are shown in table 4, where the highest accuracy was achieved with the use of ANN and GMM. Additionally, table 4 shows the sensitivity of each algorithm to each of the faults. Notice that SVM and ANN are more sensitive to Faults 1-5, corresponding to faulty sensors, while the GMM algorithm has excellent sensitivity for faults 6, 7 corresponding to faults in the actuators. Moreover, the KNN algorithm exhibited the worst performance.

## SIMULATION TESTS AND RESULTS

Given the performance results obtained by each model, the FDI system was built for Level 1 using the support vector machines, and for Level 2 GMM. Table 5 lists the average detection time and missed faults over 50 test runs. Fault 6, which represents a pitch actuator, has the longest detection times because this fault is simulated in regions 2 and 3. The problem is that the control scheme uses pitch angle as a control variable only in region 3, hence only in this region will it be possible to detect this fault. In general, it is very difficult to detect faults associate with pitch movement in region 2 because blade pitch angles are assumed to be constant. Faults 1 to 4 have acceptable detection

times. Fault 7 presents long detections times since it is an incipient fault, which can only be detected (in this case) until the movement of pitch 3 is considerably degraded.

## CONCLUSIONS

This paper presents a multi-level DD FDI system architecture that provides enhanced performance compared to single level DD FDI. In each level, four different classification strategies (SVM, ANN, KNN and GMM) were implemented and tested, where SVM and ANN showed higher accuracy for binary classification (Level 1); for multi class classification (Level 2) GMM based FDI exhibited the best performance, in particular for actuator fault detection. The FDI system built is capable of recognizing all failures selected with acceptable detection times. One of the biggest challenges in FDI is reducing the number of false alarms. Although dividing the problem into two levels and adding a FCTW to the first level of detection achieved an important reduction in false alarms, this is still an open research problem.

Data Driven classification techniques have the advantage of not needing a mathematical model of the system, and rather tries to model data behavior. Although this means that less information about the system is needed, in many cases appropriate attribute sets may only be obtained through an *ad hoc*

**Table 5.** Results Multi Level FDI Scheme.

Fault No.	Detection Time			False Alarm / Missed Detection		
	Average	Max	Min	FA	FA Clear Time	Missed
1	4.2Ts	7Ts	3Ts	2	3,5 Ts	0
2	4.3Ts	8Ts	3Ts	4	3,4 Ts	0
3	4.2 Ts	7Ts	3Ts	1	3 Ts	0
4	17.2 Ts	24Ts	8Ts	3	4.5 Ts	0
5	5.2 Ts	9Ts	3Ts	3	4.5 Ts	0
6	5125 Ts	5231Ts	5055Ts	3	3.4 Ts	0
7	2325.4 Ts	2725Ts	1627Ts	5	4.2 Ts	0

**Source:** own work.

process that requires some expert knowledge about the process. This may be particularly useful if there is some historical knowledge regarding probability of occurrence of each fault and their cause.

One of the drawbacks in this experimentation is that we had a large data set that allowed us to construct the training set in a rather simple manner. In practical applications this may generally not be the case, and the quality of the classifier may be affected. Although there are different techniques that can be used to deal with small or unbalanced data sets (Chawla, 2005), it is still important to have suitable historical data in order to be able to achieve the desired results.

## REFERENCES

- Alpaydin, E. (2004). *Introduction to Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.
- Bergantino, C. F., & S, L. (2009). FaultBuster: data driven fault detection and diagnosis for industrial systems.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer-Verlag New York, Inc.
- Chawla, N. (2005). Data Mining for Imbalanced Datasets: An Overview. En O. Maimon, & L. Rokach (eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 853-867). Springer US.
- Chen, W., & Saif, M. (June 2006). Fault detection and isolation based on novel unknown input observer design (6 pp.)
- Duda, R.; Hart, P., & Stork, D. (2001). *Pattern Classification*. Wiley.
- GWEC Global Energy Council (2011). *Global wind report, Annual market update 2010*.
- Kusiak, A., & Shah, S. (2006). Data-mining-based system for prediction of water chemistry faults. *IEEE Transactions on Industrial Electronics*, 53(2), 593-603.
- Lipnickas, A. (2006). Two Stage Neural Networks Based Classifier System for Fault Diagnosis. En V. Palade, L.; Jain, & C. D. Bocaniala (eds.), *Computational Intelligence in Fault Diagnosis* (pp. 209-230). Springer London.
- Odgaard, K. M. (2009). Fault Tolerant Control of Wind Turbines a benchmark mode. *In Proceedings of Fault Detection, Supervision and Safety of Technical Processes*.
- Scholkopf, B., & Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.
- Sylvain, T., & Kobi, A. (2010). Fault detection of univariate non-Gaussian data with Bayesian network.
- Tan, P.-N.; Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining* (1st ed.), Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
- Verhaegen, M. (2008). Fault detection of large scale wind turbine systems: A mixed H/H index observer approach. *2008 16th Mediterranean Conference on Control and Automation (M)*, 1675-1680.
- Walford, C. A. (2006). Wind Turbine Reliability: Understanding and Minimizing Wind Turbine Operation and Maintenance Costs. *Energy*(March), SAND2006-1100.
- Zhou, Y. (2004). *Data Driven Process Monitoring Based on Neural Networks and Classification Trees*. Master's thesis, Texas A&M University.

