

# Environmental Sensor Drift Correction using Wavelet-mFCM-FIS Architecture: An Unsupervised Machine Learning Approach

Ritaban Dutta

Intelligent Sensing and Systems Laboratory,  
CSIRO, Hobart, TAS 7001, Australia

## ABSTRACT

A novel hybrid sensor informatics architecture based on Discrete Wavelet Transform (DWT), Fuzzy logic based clustering (FCM) and Fuzzy Rule based Inference system (FIS) has been investigated and proposed to estimate dynamic generic sensor drift during physical sampling. DWT has been used for sensor pre-processing; data dimension reduction and feature extraction from sensor time series, where as DWT-FCM based approach has been used to estimate the cumulative drift in our sensory system. An intelligent generalized algorithm using multiple FCM maps (m-FCM) has been proposed to implement the drift correction on any future unknown data sets. In the next stage a newly proposed DWT-FCM-FIS architecture has been tested on multisensory environmental data sets covering three consecutive years (2009 – 11) to predict probable month (environmental time of the year) with up to 93% accuracy. This novel hybrid data processing architecture have been implemented and tested upon a real time estuary sensory platform of temperature, conductivity and salinity sensors that have been deployed to monitor the Derwent Estuary in Hobart, Australia. This newly proposed approach could be an adaptive solution to tackle drifts in the sensor networks and improve overall monitoring quality.

## General Terms

Sensor Drift Correction, Oceanic Data Analysis, Pattern Recognition.

## Keywords

Sensor Drift, Drift Area, Discrete Wavelet Transform, Fuzzy C means, Fuzzy Inference system.

## 1. INTRODUCTION

Environmental drift of ocean sensors is a dynamic process caused by physical changes in the sensors and the chemical background, which gives an unstable signal over the time. Sensory data are usually affected by variations which are partly due to the non-ideal of sensor response and partly due to the effects of the environment on the chemical sensory component. It is widely acknowledged that the sensor networks are expensive and there are significant risks associated with the deployment and maintenance. The main monitoring area for this study was a coastal region that covered an approximate area of 100x50km, so a very practical approach was taken in the CSIRO's Tasmanian Marine Analysis Network project (Tas-MAN) to the implementation of a marine sensor network. A low-cost sensors network platform was deployed. The low-cost sensors and platform do not provide the high quality observations and diversity of phenomena, but by sacrificing accuracy at one geographic point improved representation was achieved for a larger area. But greater uncertainty, low resolution in sensor data quality

and significant chances of sensor drift created a unique opportunity to develop an adaptive sensor drift estimation framework for the ocean sensor network. Successful development of this type of framework could potentially enhance the drift correction capabilities which in turn will enhance the reliability of the sensor network and reduce the maintenance worries. In this paper novel hybrid sensor informatics architecture based on Discrete Wavelet Transform (DWT) and multiple Fuzzy C Means clustering (m-FCM) has been investigated and proposed to estimate dynamic sensor drift and potential drift correction [1-4].

In this paper novel hybrid sensor informatics architecture based on Discrete Wavelet Transform (DWT), Fuzzy logic based clustering (FCM) and Fuzzy Inference system (FIS) has been investigated and proposed to estimate dynamic sensor drift and potential drift correction. DWT has been used for sensor pre-processing, data dimension reduction and unsupervised feature extraction from sensor time series, where as FCM and FIS has been used to develop the hybrid drift estimation incorporated predictive model [5-8].

## 2. DATA AND SYSTEM

This particular study is based on six Tyco Greenspan EC250 sensors [9] which are being used for measuring temperature, conductivity and salinity of the sea water. Figure 1 depicts a typical sensor node deployed into ocean water for measurements.

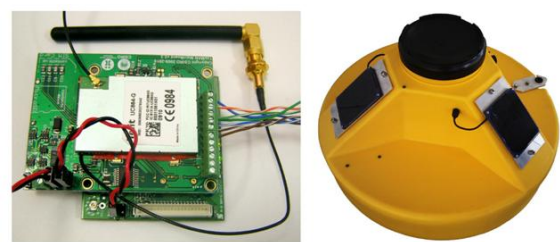
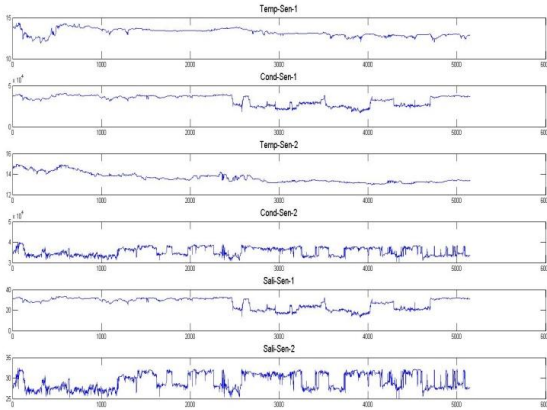


Fig 1: Tyco Greenspan EC250 sensors

Two sensors are dedicated for contentious measurement of each of the parameters. Data were gathered using web servers. In the Figure 2 typical sensor responses are shown which were captured during May 2009. The TasMAN system's integrated design avoids the range limitation and brittleness of mesh networking by making cellular communications affordable on every sensor node, and significantly improving data throughput.



**Fig 2: Typical monthly Sensor observations during May 2009.**

The whole study was based on real sensory observations during three separate time periods namely {Set 1 = Jan 2009 – July 2009}, {Set 2 = Jan 2010 – July 2010}, and {Set 3 = Jan 2011 – July 2011}. So effectively each set had six months of data representing same time of the year with same length period. In this paper, data modelling was done on a monthly basis using Wavelet – FCM – FIS approach for the final FIS to predict the probable month from the multi-dimensional sensor data. On the other hand Wavelet – FCM approach was taken to measure sensory drift over the years. Chronological sections of this paper are written in a way that they would follow the exact flow of our data analysis and predictions. Firstly data sets were pre-processed, normalized, and the representative features were calculated using wavelet analysis. Then the pre-processed features were used to model the relationship between the input variables (six sensory responses) and the output variable (actual month) initially by clustering the data. The cluster centers were used as a basis to define a FIS which was subsequently used to explore and understand environmental patterns over the years. Set 1 data is used in the Wavelet – FCM – FIS based initial water data understanding and modelling where as Set 2 and Set 3 are used to measure the sensory drift over the years and to test prediction performance of the system [10-13].

## 2.1 Data Normalisation

In this study, normalisation model in the Equation 2 have been used to make the responses linear, to increase the relative contribution of sensors and the overall dynamic range. Sensors were normalised individually based on the maximum and minimum values for that particular time period.

$$R_{ij} = \frac{(x_{ij} - x_i^{\min})}{(x_i^{\max} - x_i^{\min})} \quad (2)$$

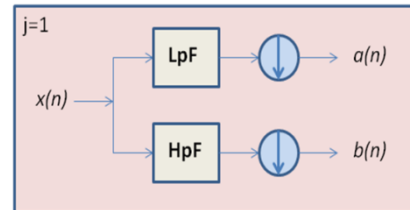
## 2.2 Discrete Wavelet Transform: Feature Extraction

In this paper Discrete Wavelet Transform (DWT) is proposed as a time-series feature extraction algorithm for automatically choosing feature and reducing dimensionality for clustering. DWT transforms the time series from time domain into time-frequency domain. It has been identified that DWT could be an ideal feature extraction technique with the ability to efficiently reduce the data into a lower dimensional model, while preserving the properties of the original data. In practice, however, information is lost as the dimensionality is reduced. It is therefore desirable to formulate a method that

reduces the dimensionality efficiently, while preserving as much information from the original data as possible.

Due to difficult oceanic physical scenario for data gathering it is evident that total number of sensor observations recorded each month vary from month to month. So for this study total number of representative sensory observations for each month has been selected based on number of stages of wavelet analysis.

So final target time series length  $n$  for individual sensor  $i$  was selected in a way where  $n$  is divisible by  $2^j$ , and  $n \geq \text{length of } \{LpF, HpF\} * 2^{j-1}$  where  $j$  is number of stages of wavelet analysis.



**Fig 3: Single stage schematic representation of Wavelet Analysis.**

The low pass filter is denoted by LpF and the high pass filter by HpF. This is the first step towards overall dimension reduction process for further feature based data clustering.

In each stages the analysis filter bank decomposes the input signal  $x(n)$  into two sub band signals,  $a(n)$  and  $b(n)$ . The signal  $a(n)$  represents the low frequency part and the signal  $b(n)$  represents the high frequency part of  $x(n)$ . As shown in the figure, the output of each filter is then been down-sampled by 2 to obtain the two sub band signals,  $a(n)$  and  $b(n)$ .  $n$  length signal from individual sensor was decomposed into coarser resolutions using a simple hierarchical scheme and the schematic diagram of wavelet composition is shown in Figure 3. In this paper four-level wavelet transform was used for dimension reduction and feature extraction. In order to find the optimum number of wavelet level required for best possible representation of features from the data a feedback optimization was applied based on feature based component data clustering (Figure 4).

Different levels of wavelet were applied to reduce dimension and extract features until the first three components (in principal component analysis (PCA)) were covering 99.99% of information variance among the data sets. It was found that four-level wavelet transform was best suited for this study where first two PC covered more than 99% of information variation [14]. Individual stage of decomposition process provided detailed  $j+1$  coefficients where as  $j^{th}$  coefficient after the final stage of wavelet was eventually used as representative feature of that class or month in this case [14-17].

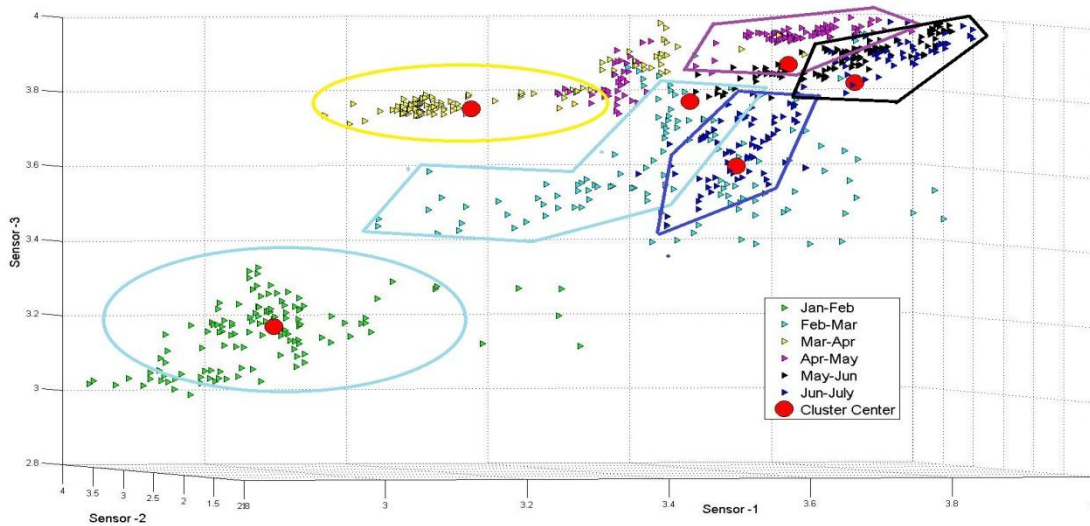


Fig 4: Clustering using PC-FCM algorithm to identify six separable monthly data

### 3. FUZZY C MEANS

Fuzzy C means (FCM) is an unsupervised method. Clustering essentially deals with the task of splitting a set of patterns into a number of more-or-less homogeneous classes (clusters) with respect to a suitable similarity measure such that the patterns belonging to any one of the clusters are similar and the patterns of different clusters are as dissimilar as possible. The similarity measure used has an important effect on the clustering results since it indicates which mathematical properties of the data set should be used and in what way they should be used in order to identify the clusters [18-22]. FCM clustering provides partitioning results with additional information supplied by the cluster membership values indicating the degrees of belongingness, where C is the total number of clusters.

The objective function of the FCM algorithm takes the form of

$$J(u_{ij}, v_k) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2, \quad m > 1, \quad (3)$$

where m is called the exponential weight which influences the degree of fuzziness of the membership (partition) matrix. The FCM algorithm provides an iterative approach to approximating the minimum of the objective function starting from a given position. This iterative procedure minimizes the objective function in Equation 3 [14, 20, 23]. It was tested and observed that pre-processed and extracted feature base did not have local convergence, so it won't impact drift estimation and drift correction results.

#### 3.1 FCM Clustering: Drift estimation

FCM clustering provides a probable center of the clusters along with the additional information supplied by the cluster membership values indicating different degrees of belongingness. These cluster centers are the direct feature representation of a cluster which is representing a particular class of data. In an unsupervised fashion, these cluster centers could be estimated among the whole data set and particular classes or groups inside the data could be classified.

In Figure 4 example results from the FCM clustering on Set 1 are presented. Cluster centers are marked with red dots where as six different clusters are marked with different colours. FCM algorithm has predicted the positions of six most

important clusters from the whole data set containing interesting knowledge about the data which could be used to estimate amount of sensor drift from year to year. Ideally if the sensors have no drift over the years and operate consistently then shift of movement of the cluster centers would be very limited. In reality that is not the case and the response from sensors contain significant amount of drift. It is evident from this figure that clusters which are representing 'Jan-Feb', 'Mar-Apr' and 'Jun-July' are very clearly separable. The other three clusters representing intermediate transition months have overlapping boundaries with clusters. In order to estimate sensor drift more accurately it is important to distinguish noticeable differences in cluster centers due to natural variation in temp and conductivity phenomena between the years from the actual year to year sensor drift. After studying the three yearly data independently it was evident that if we re-cluster the feature sets into three more consistent clusters we can represent the actual aspects of sensor drift more, which minimized the effect of sensory natural variations. So if we model the whole data set into three major cluster regions, the problem could be easily represented and the amount of drift could be calculated more consistently with high accuracy, as weather wise it is quite practical to divide the period from January to July into three different classes.

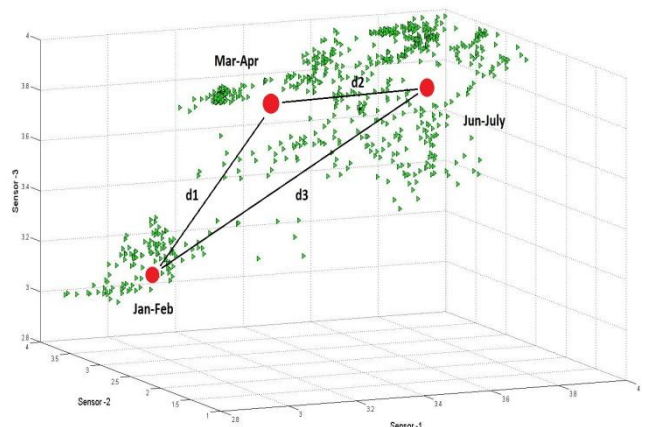


Fig 5: FCM based re-clustering and development of drift calculation algorithm.



So with that plan in place the whole data Set 1 was re-clustered into three clusters. Figure 5 shows the three new cluster centers and distances between two clusters are marked as  $\{d_1+ d_2+ d_3\}$ . One of the most important aspect of this paper is that we wanted to define a new way to estimate drift in the sensory system numerically, so in order to parameterize the amount of drift, the area of the triangle formed by three cluster centers has been calculated as *Drift Area<sub>i</sub>* using Heron's formula (Equation 4).

$$Drift Area_i = \sqrt{(s*(s-d_1))*(s-d_2)*(s-d_3)} \quad (4)$$

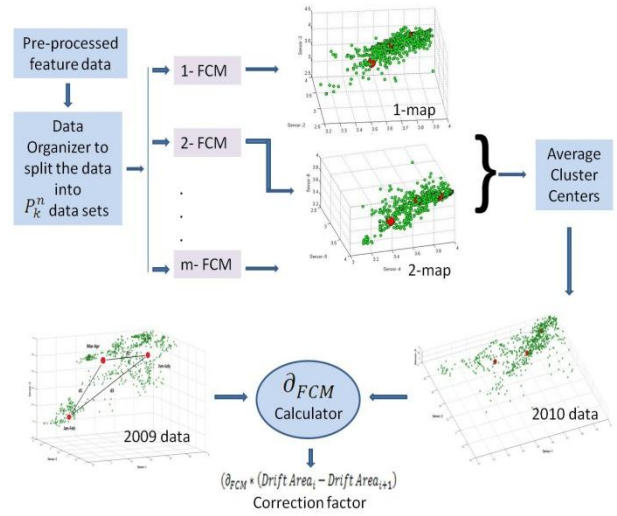
where,  $s = \frac{d_1 + d_2 + d_3}{2}$ ,  $i = 2009, 2010, \dots$

Next both Set 2 and Set 3 representing same time period from 2010 and 2011 were analysed in the same way as Set 1. These individual year wise clustering enabled us to estimate  $Drift Area_{2010} = 0.1831$  and  $Drift Area_{2011} = 0.1169$  and compare those against  $Drift Area_{2009} = 0.2938$  to measure cumulative drift in our sensory system over last three years. Difference between two drift areas is the estimation of amount of drift for that particular one year period. The drift areas from 2010 and 2011 were compared with the drift area from 2009 as sensors were deployed in 2008, so it was expected that sensor performance and sensibility were much higher in 2009 than in 2010-11. From the clustering and drift estimation it is evident that these three clusters are always separable from each other on the basis of individual analysis of Set 1, Set 2 and Set 3 data, but over the years  $Drift Area_i$  is getting smaller which means separability among the clusters are getting narrower, or in other words the clusters are becoming more closer to each other over the years. So indirectly this analysis is giving us an exact numerical estimation of gradual reduction of sensibility among the sensors responses.

### 3.2 m-FCM based Drift Correction

Once we have a FCM based clustering map from an unknown data set, then the labeling phase follows the unsupervised one, in order to classify data based on the Euclidean distance. However, in the context of the oceanic sensors the use of a single map often becomes useless due to drift. If a sensor is not behaving in the expected manner, it would not map the latest probability distribution of the new physical environmental scenario. In other words, if a cluster moves to a new position, it is not obvious that all the feature points belonging to that cluster will be updated. This behaviour could give rise to confusion, since in the middle of a cluster there could be a feature point that belongs to another cluster and it has not been represented correctly for a very long time [4, 14].

For this reason a novel approach has been proposed to incorporate the drift correction based on the drift estimation in previous section 3.1 and combining that with multiple FCM architecture (m-FCM). Figure 6 shows the m-FCM architecture that has been developed for drift correction. A multi FCM map has been developed that assures the self-adjustment process to all the feature points in the local map, and autonomous adaptation to new situations. This approach preserves the self-organization paradigm by considering as many maps as the number of monthly environmental scenarios, in order to accomplish the classification task.



**Fig 6: m-FCM based drift correction factor estimator algorithm**

In the unsupervised analysis stage multiple data sets have been created based on all combinations of three sensors out of possible six sensors. That was done using permutation as shown in the Equation 5.

$$P_k^n = n * (n-1) * (n-2) \dots (n-k+1) \quad (5)$$

$$P_k^n = 0 \text{ when } k > n, \text{ else } P_k^n = n! / (n - k)!$$

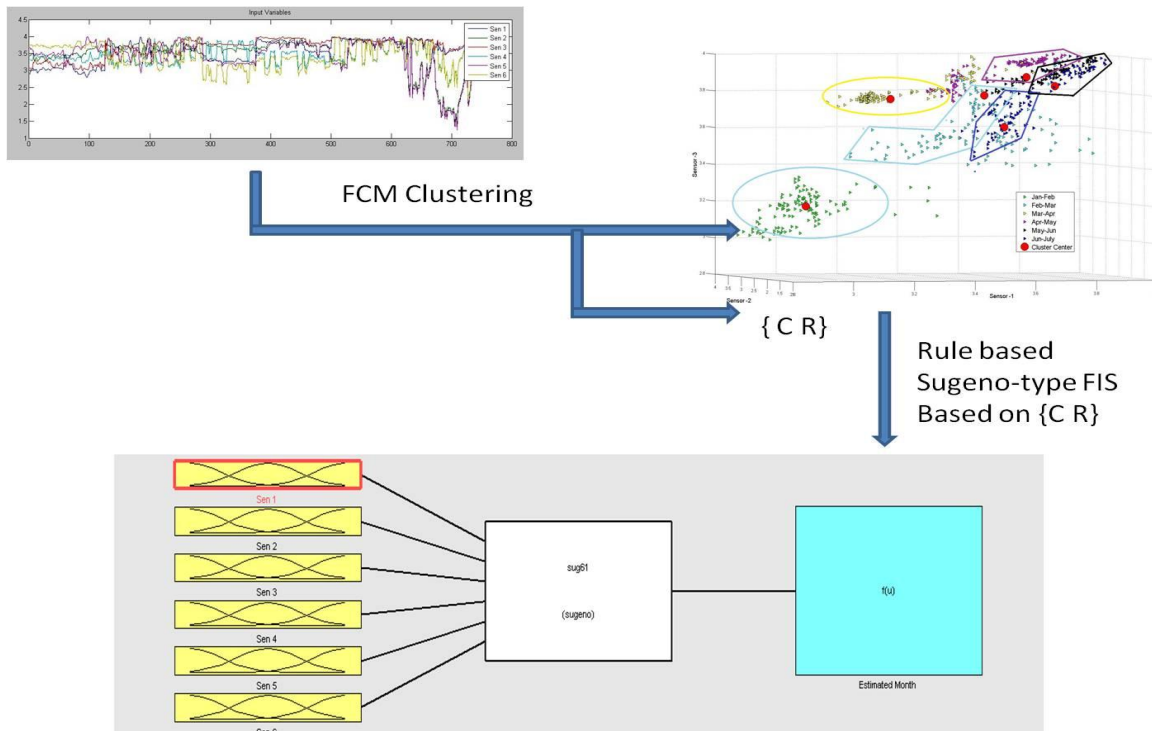
Each sensory data combination representing same data set was clustered separately using FCM algorithm to produce  $P_k^n$  independent maps with independent sets of three possible cluster centers. All three data sets were pre-processed and clustered into three clusters to estimate the probable relative cluster positions representing the same period of the year. Although the whole process is completely unsupervised but in order to track the movements of the clusters this independent and unbiased way, this type of supervision of the data was used. In the next stage all relatively calculated cluster centers (which are effectively representing same data cluster for the same period of the year) were combined to calculate an average position of that particular cluster center. These newly positioned cluster center was then compared with previous year's cluster center representing same data period to estimate difference in physical cluster position. This is done based on Euclidean distance measurement and is called in this paper as  $\delta_{FCM}$ . Three different cluster centers were then associated with three different values of  $\delta_{FCM}$ . In Figure 6 different stages of this algorithm has been illustrated. Main steps of the proposed algorithm are as follows:

**Step 1:**  $P_k^n$  cluster centers from  $P_k^n$  independent FCM for  $k^{\text{th}}$  cluster

**Step 2:** Average cluster center for  $k^{\text{th}}$  cluster

**Step 3:**  $\delta_{FCM}$  for  $k^{\text{th}}$  cluster

**Step 4:** Drift corrected feature point for  $k^{\text{th}}$  cluster =  
feature point for  $k^{\text{th}}$  cluster  $\pm (\delta_{FCM} * (Drift Area_i - Drift Area_{i+1}))$



**Fig 7: Representation for Sugeno-type FIS modelling based on FCM clustering**

To develop this estimation algorithm it was considered that if in reality all the associated sensors might have different drift factors, but as we have used a pre-processed projection study based on “least correlated sensor combination” for feature extraction and FCM clustering, the effect of individual sensor drift was estimated as a part of overall drift estimation between two years. Finally a separate clustering was performed on the three least correlated sensors for that year to capture the maximum variance among the feature data points; and then all data points in an individual cluster were corrected with the corresponding

$$(\delta_{FCM} * (Drift\ Area_i - Drift\ Area_{i+1}))$$

product value associated with that cluster’s center. This newly defined product was introduced to generalize, estimate and correct overall sensory drift. With this correction, new feature data points were achieved for that particular year. In the section 5 a comparative study on FIS based prediction has been performed between ‘original feature data points’ and ‘corrected feature data points’ in order to evaluate this physical drift correction mechanism.

#### 4. FCM-FIS DEVELOPMENT: DATA BEHAVIOURAL MODELLING

In this section we attempt to understand the relationship between the responses generated from the sea water sensory system (consists of six sensors) and the month predictability of the system. Hereon, the sensor responses will be addressed as inputs and the month predicted or labeled will be addressed as output. Hence our problem has six input variables (six water quality monitoring sensors) and one output variable (predicted month). In order to model the relationship between the input variables and the output variable, clustering of the data was the first step using FCM based subtractive clustering. The cluster centers will form a basis to define a FIS which can then be used to explore and understand monthly environmental patterns [19, 22]. Clustering can be a very

effective technique to identify natural groupings in data from a large data set, thereby allowing concise representation of relationships embedded in the data. In this example, clustering allows us to group monthly sensory patterns into broad categories hence allowing for easier data understanding. Cluster representations from the Set 1 (based on 2009 data) in Figure 5 were the foundation to model the FIS.

Fuzzy logic is an effective paradigm to handle imprecision. It can be used to take fuzzy or approximate observations for inputs and yet arrive at crisp and precise values for outputs. Also, the FIS is a simple and rational way to build systems without using complex analytical equations. In our example, fuzzy logic will be employed to capture the broad categories identified during clustering into a FIS. The FIS will then act as a model that will reflect the relationship between sensor observations and predicted month of the year. Clustering and fuzzy logic together provide a simple yet powerful means to model the relationship between sensor and physical oceanic parameters that we want to study. Figure 7 represents the schematic diagram of the whole Sugeno-type FCM-FIS system that has been used in this study. FCM based subtractive clustering estimates the cluster centers in a set of data by using the subtractive clustering method. The subtractive clustering method assumes each data point is a potential cluster center and calculates a measure of the likelihood that each data point would define the cluster center, based on the density of surrounding data points. The cluster’s radius of influence in the input space was set to 3.5 for justification against significant amount of overlapping among the clusters.

The variable C in Figure 7 holds all the centers of the clusters that have been identified by FCM based subtractive clustering. Each row of C contains the position of a cluster. In this case, C has 6 rows representing 6 clusters with 6 columns representing the positions of the clusters in each dimension. The variable S in Figure 7 contains the sigma values that

specify the range of influence of a cluster center in each of the data dimensions. All cluster centers share the same set of sigma values. S in this case has 6 columns representing the influence of the cluster centers on each of the 6 dimensions.

Given separate sets of input and output data, Sugeno-type FIS structure was generated using results from fuzzy clustering by extracting a set of rules that models the data behaviour. The rule extraction method first uses the FCM algorithms to determine the number of rules and membership functions for the antecedents and consequents. The number of clusters determines the number of rules and membership functions in the generated FIS. As can be seen in Figure 7, the FIS has 6 inputs and 1 output with the inputs mapped to the outputs through a rule base (white box in the figure).

It is important to understand how the cluster centers (C) are related with the rule base generated in the FIS. Figure 8(a)-8(f) represent the individual rule base scenarios related to individual sensor inputs. Notice that all the inputs and outputs have exactly 6 membership functions. The 6 membership functions represent the 6 clusters that were identified by FCM clustering. Membership functions are called as {in1cluster1, in1cluster2, in1cluster3, in1cluster4, in1cluster5, in1cluster6}. The membership function type used in this study is a Gaussian type membership function (MATLAB based function described in Equation 6)

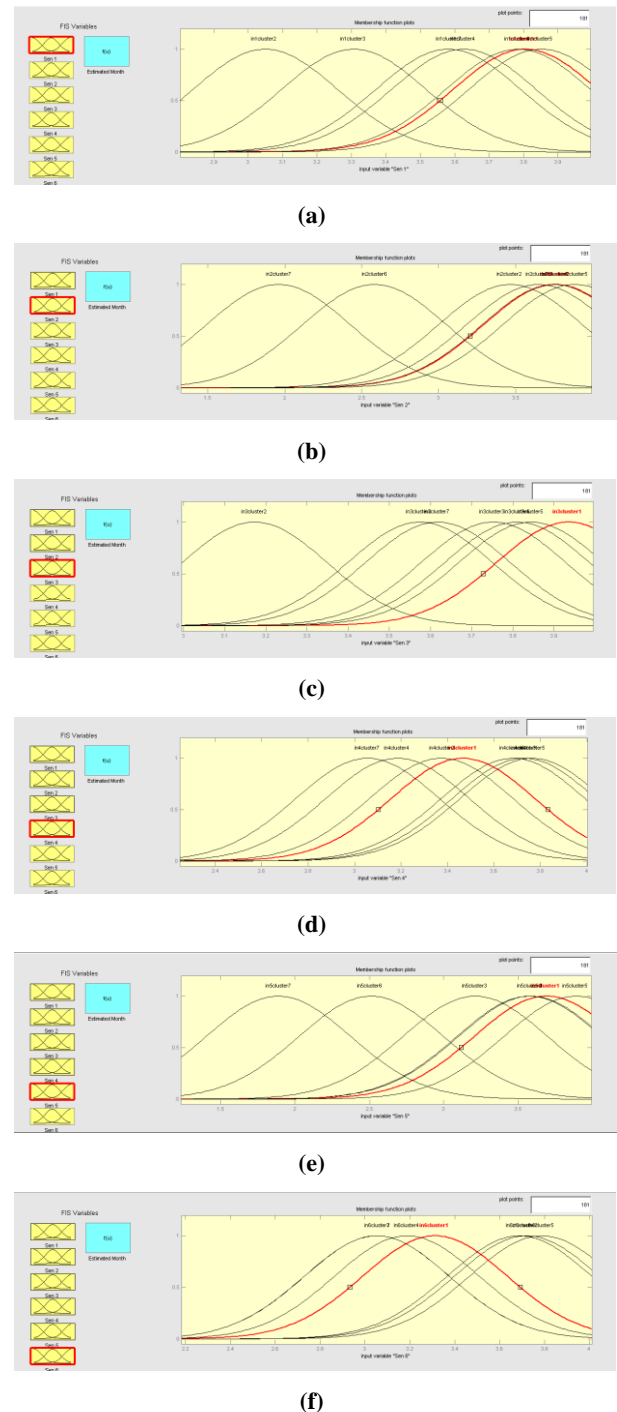
$$f(x; \delta, c) = e^{-\frac{(x-c)^2}{2\sigma^2}} \quad (6)$$

and the parameters of the membership function are defined by the results (C and R) from the clustering exercise; where corresponding sigma values represents the spread coefficient of the Gaussian curve and cluster centers represent the center of the Gaussian curves. Cluster1 captures the position and influence of the first cluster for the input variable sensor 1 and so on. Similarly, the position and influence of the other 5 clusters for the input variable sensor 1 are captured by the other five membership functions in cluster2, in1cluster3, in1cluster4, in1cluster5 and in1cluster6. The rest of the 5 inputs follow the exact pattern mimicking the position and influence of the 6 clusters along their respective dimensions in the dataset. As seen, there are exactly six rules. Each rule attempts to map a cluster in the input space to a cluster in the output space. There are exactly six rules of which each rule attempts to map a cluster in the input space to a cluster in the output space. The first rule can be explained simply as follows.

If an unknown data point closer to the first cluster, or in other words having strong membership to the first cluster, is fed as input to FIS then rule1 will fire with more strength than the other two rules. Similarly, an input with strong membership to the second cluster will fire the second rule will with more strength than the other two rules and so on. If the inputs to the FIS, Sensor 1, Sensor 2, Sensor 3, Sensor 4, Sensor 5 and Sensor 6, strongly belong to their respective cluster1 membership functions then the output, estimated or predicted month, must strongly belong to its cluster1 membership function. The (1) at the end of the rule is to indicate that the rule has a weight or an importance of "1". Weights can take any value between 0 and 1. Rules with lesser weights will count for less in the final output. The significance of the rule is that it clearly maps cluster 1 in the input space to cluster 1 in the output space. Similarly this same rule structure applies to other 5 rules.

Finally a MATLAB GUI based FIS system were developed to visualize and test the performance of FIS system which is shown in Figure 9. Now, having built the FCM-FIS system, if

we want to understand the weather wise time of the year or month for a particular six sensor based demographic setup, this tool will help to simulate the FIS response for the input combination of the choice.



**Fig 8: Example shape of the Membership Function Plots for Rule base in Sugeno-type FIS, (a) Sensor 1, (b) Sensor 2, (c) Sensor 3, (d) Sensor 4, (e) Sensor 5, and (f) Sensor 6.**

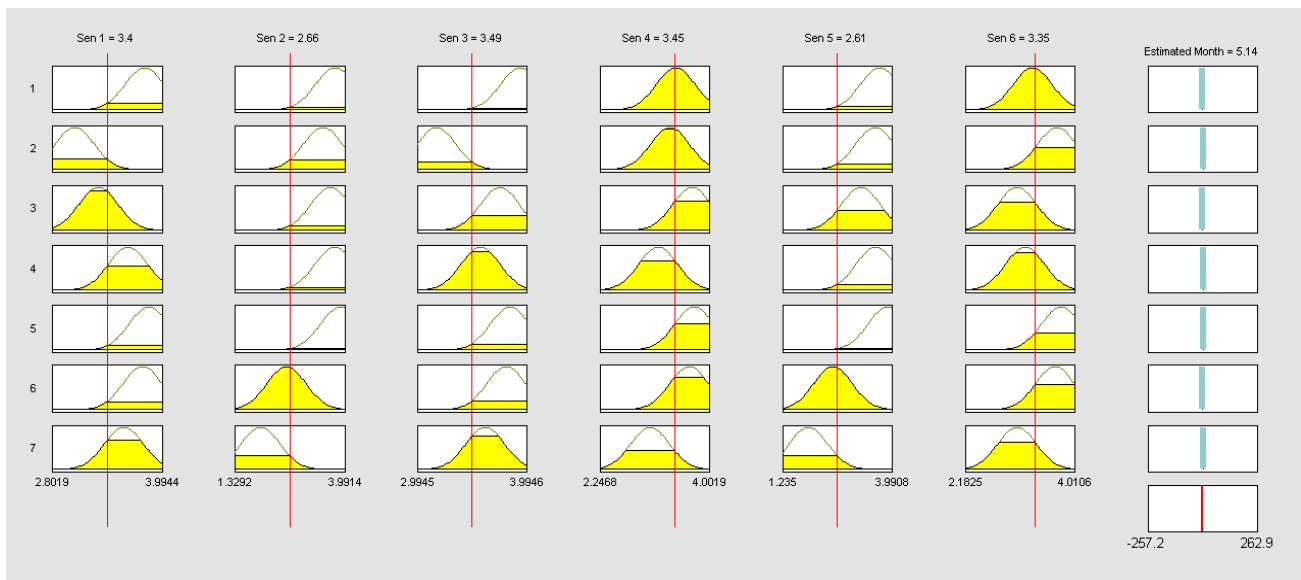
In the next section we will summarize the predictability performance of the newly designed FIS system. As described previously, the whole FIS system was modelled using 2009 data Set 1, whereas all performance evaluations were done using Set 2 (2010) and Set 3 (2011) to make the overall study a more accurate and reliable one [20, 22]. Table 1 includes values of the estimated cluster centers from the data Set 1.

**Table 1: The estimated cluster centers from the data Set 1.**

	Dimension 1	Dimension 2	Dimension 3	Dimension 4	Dimension 5	Dimension 6
Cluster 1	3.0518	3.8064	3.8029	3.3448	3.6604	3.5677
Cluster 2	3.4564	3.7948	2.6493	3.6288	3.7088	1.8559
Cluster 3	3.1771	3.8688	3.8365	3.7559	3.7720	3.6249
Cluster 4	3.3735	3.6496	3.6350	3.6670	3.3011	2.9111
Cluster 5	3.5776	3.7695	2.5723	3.2556	3.6354	1.7839
Cluster 6	3.7207	3.6149	3.6273	3.0923	3.2149	2.8920
Sigma Values (S)	1.1765	0.3117	1.6555	6.6139	2.8931	1.9395

**Table 2: The estimated cluster centers from the data Set 1.**

	Results from FIS System without Drift Correction (% Accuracy)	Results from FIS System with Drift Correction (% Accuracy)
DATA SET 2	79.92%	89.7%
DATA SET 3	81.63%	92.89%



**Fig 9: Graphical Representation of FIS System and prediction capabilities against a particular set of input from all six sensors.**

## 5. FIS SYSTEM PERFORMANCE

To evaluate the FIS system a separate data pre-processing and data feeding mechanism were developed which was independent from the original FIS system. FIS performance evaluation was done on the basis of data Set 2 and Set 3. First of all data were pre-processed and then wavelet analysis was used to extract key features from the data base. In the next stage those features were treated as input to our FIS system. Typical combination of an input was a combined processed observation from the six oceanic sensors. Effectively the individual observation values from six sensors were moving the sliding bars on the FIS system. Each time a new output was recorded as a probable month against a sensor input combination. Final month outputs were rounded to figure out exact predictions. Results were quite encouraging as overall prediction performance for Set 2 was 79.92% whereas for Set 3 it was 81.63% accurate. These were achieved using uncorrected feature data points as input to the FIS system. In the next stage of testing drift corrected feature data points were used as input to the FIS system and the overall prediction performance for Set 2 was 89.7% whereas for Set 3 it was 92.89% accurate. It was easily evident that m-FCM

based drift correction mechanism heavily impacted on the FIS system performance in a positive direction (Table 2).

## 6. CONCLUSION

In this paper novel hybrid sensor informatics architecture based on Discrete Wavelet Transform (DWT), Fuzzy logic based clustering (FCM) and Adaptive Fuzzy Rule based Inference system (FIS) has been investigated and proposed to estimate dynamic sensor drift. DWT has been used for sensor pre-processing; data dimension reduction and feature extraction from sensor time series, where as DWT-FCM based approach has been used to estimate the cumulative drift in our sensory system. A novel algorithm using multiple FCM map (m-FCM) has been proposed to implement the drift correction on any future data. In the next stage a newly proposed DWT-FCM-FIS architecture has been tested on multisensory environmental data sets covering three consecutive years (2009 – 2011) to predict probable month (environmental time of the year) with up to 93% accuracy.

The TasMAN project's holistic approach to low-cost marine monitoring is producing innovations in many areas of coastal

observing. It is evident from the results of this study that combination of proposed algorithms along with the TasMAN project technology will improve the feasibility of coastal monitoring on a much larger scale, particularly tackling the long term sensor drifts. This paper presents a hybrid unsupervised system which could enable government, community and industrial organizations with small budgets to conduct their own monitoring programs.

This study has produced a strong foundation of an unsupervised approach to minimize drift in the estuary sensor network. This work was mainly focused to motivate low cost maintenance approach by keeping in line with our "low cost sensing" theme where this novel approach could potentially reduce voyage/maintenance costs. Further research on different data sets using these novel approaches and knowledge regarding effects on the health of our waterways will encourage informed positive changes.

## 7. ACKNOWLEDGMENTS

The author wishes to thank the Intelligent Sensing and Systems Laboratory and the Tasmanian node of the Australian Centre for Broadband Innovation. A grant from the Tasmanian Government which is administered by the Tasmanian Department of Economic Development, Tourism and the Arts helped to conduct this research. The TasMAN project sensor node hardware has been funded by the CSIRO Sensor and Sensor Networks Transformational Capability Platform (SSN-TCP). Author would like to acknowledge Dr Daniel Smith for his supports by providing literatures and web server based data gathering protocols.

## 8. REFERENCES

- [1] C. Sharman, "Tasmanian Marine Analysis Network: Portal," <http://www.csiro.au/tasman/portal>, July 2011.
- [2] Ding hui, Liu Jun-hua, Shen Zhong-ru, Drift reduction of gas sensor by wavelet and principal component analysis, *Sensors and Actuators B* 96 (2003) 354–363.
- [3] Alexander Vergaraa, Shankar Vembua, Tuba Ayhan, Margaret A. Ryan, Margie L. Homer, Ramón Huertaa, Chemical gas sensor drift compensation using classifier ensembles, *Sensors and Actuators B* 166– 167 (2012) 320– 329.
- [4] M. Zuppa, C. Distanta, P. Siciliano, K.C. Persaud, Drift counteraction with multiple self-organizing maps for an electronic nose, *Sens. Actuators B Chem.* 98 (2004) 305– 317.
- [5] Australia's Integrated Marine Observing System (IMOS), <http://imos.org.au/> (2012)
- [6] C. D'Este, A. Terhorst, G. Timms, J. McCulloch, and C. Sharman, "Adaptive marine monitoring via the sensor web," in *IEEE Conference on Intelligent Robots and Systems: Workshop on Robotics for Environmental Monitoring*, San Francisco, USA, 2011.
- [7] G. Timms, B. Howell, P. de Souza, C. D'Este, D. Smith, R. Atkins, and C. Sharman, Automated assessment and visualisation of measurement uncertainties in a low-cost marine sensor network, in *European Geographic Union General Assembly*, Vienna, Austria, 2011.
- [8] R. R. Brooks and S. S. Iyengar, *Multi-Sensor Fusion: Fundamentals and Applications with Software*, Prentice Hall PTR, Upper Saddle River, New Jersey 07458, ISBN 0-13-901653-8.
- [9] <http://www.greenspan.com.au/> (2012)
- [10] T.L. Chen, R.Z. YouA, Novel fault-tolerant sensor system for sensor drift compensation, *Sensors and Actuators A* 147 (2008) 623–632.
- [11] A. Ziyatdinov, S. Marcoe, A. Chaudry, K. Persaud, P. Caminal, A. Pereraa, Drift compensation of gas sensor array data by common principal component analysis, *Sensors and Actuators B* 146 (2010) 460–465.
- [12] A. Ray, R. Luck, An introduction to sensor signal validation in redundant measurement systems, *Control Systems Magazine IEEE* (1991) 44–49.
- [13] R Dutta, D Morgan, N Baker, J Gardner, E Hines, Identification of Staphylococcus aureus infections (MRSA, MSSA and C-NS) in hospital environment: Electronic Nose based approach, *Sensors and Actuators B: Chemical*, 109 (2), pp.355-362.
- [14] F.W. Fairman, *Linear Control Theory: The State Space Approach*, John Wiley & Sons, 1998, pp. 147–166.
- [15] R Dutta, OBJECT ORIENTED EXPERT ELECTRONIC NOSE SYSTEM, in *ENCYCLOPEDIA OF SENSORS*, Volume 7, Section 'O', Chapter 1, pp.1-54, American Scientific Publishers. *Encyclopedia of Sensors*, 10-Volume Set. Edited by Craig A. Grimes, Elizabeth C. Dickey, and Michael V. Pishko. The Pennsylvania State University, University Park, USA. FOREWORD by Professor Rudolph A. Marcus, Nobel Prize Laureate, 2006ca. 8,000 pages, Hardcover, ISBN: 1-58883-056-X.
- [16] F. Hossein-Babaei, V. Ghafarinia, Compensation for the drift-like terms caused by environmental fluctuations in the responses of chemoresistive gas sensors, *Sensors and Actuators B* 143 (2010) 641–648.
- [17] G. Hu,W. Chen, Y. Chen, D. Liu, Adaptive kalman filtering for vehicle navigation, *Journal of global position systems* 2 (1) (2003) 42–47.
- [18] B. Howell, *Tasmanian Marine Analysis Network: Web Data Service*, <http://www.csiro.au/tasman/WDS/wds?request=GetDeployment&format=xml>, (2011).
- [19] *Wikipedia Website Main Page* (2012)
- [20] J.J. Slotine, W. Li, *Applied Nonlinear Control*, Prentice Hall, 1991, pp. 40–97.
- [21] C. T. Lin and C. S. G. Lee, *Neural Fuzzy Systems A Neuro-Fuzzy Synergism to Intelligent Systems*, Prentice Hall, Upper Saddle River, NJ 07458, ISBN 0-13-235169-2, 1996.
- [22] H.K. Khalil, *Nonlinear Systems*, third edition, Prentice Hall, 2000, pp. 111–181.
- [23] [www.mathworks.com](http://www.mathworks.com) (2012).