Actuator Fault-Detection for Autonomous Underwater Vehicles Using Unsupervised Learning

Matt Kemp[1], and Ben Raanan[2]

[1,2] *Monterey Bay Aquarium Research Institute, Moss Landing, CA, 95039, USA*
mkemp@mbari.org
byraanan@mbari.org

ABSTRACT

Many Autonomous Underwater Vehicles (AUV) have high rates of false-alarms because their health management relies on user-generated rules. We suggest that the high false-alarm rate could be substantially lowered if fault-detection were based on actual actuator performance instead of heuristics. We collected data on a critical AUV actuator, a mass-shifter, in order to develop an unsupervised fault detector. We found that a small number of features were sufficient to detect known and novel faults with a high probability of detection and a low false alarm rate. We also found that n-point false-alarm reduction schemes performed poorly due to correlation during actuator startup.

1. INTRODUCTION

Autonomous Underwater Vehicles (AUV) are regularly used by military, oil & gas, and science customers, yet despite wide operational adoption their reliability remains low. Griffiths et al. (Griffiths, Millard, McPhail, Stevenson, & Challenor, 2003) analyzed 4 years of reliability of the Autosub AUV and found a mean time-to-failure of order 10h; Brito et al. (Brito, Smeed, & Griffiths, 2014) examined user-generated reliability data on the Slocum gliders and found 40% failure rates; Brito (Brito, 2015) examined the reliability of the Autosub-LR and found 20h mean time-to-failure.

Available AUV reliability data does not discriminate between actual faults and false alarms. A recent unpublished report based on 10,000 hours of operational data on Tethys-class AUVs suggested that 95% of the faults were false alarms (Bellingham, 2014); of those, 60% originated with the actuators, 35% with the control software, and 5% with the sensors. This finding suggests that reliability could be improved by as much as 20X by reducing the rate of false alarms.

One of the reasons for this high rate of false alarms is the near-universal dependence on user-generated health-management rules. This approach is expedient but it is error-prone as it relies on an incomplete picture of the vehicle's state and context.

A second reason is the relative rarity of faults: testing at 1Hz for example when faults occur every 10 hours would require a detector capable of handling events with a 3x10-5 probability of occurrence. Because faults are rare events, there usually is not enough data to train a classifier.

An alternative to supervised learning is unsupervised learning, where data groupings are found from regularities in the data instead of user-generated labels. A common instance of unsupervised learning is an anomaly detector, where the detector is trained directly from the nominal data and its performance is adjusted using a small amount of fault data.

Raanan et al. recently used topic modeling, a mixed-membership Bayesian unsupervised learning technique, to detect vertical plane failures (Raanan et al., 2016, 2017). Using no labels, the algorithm grouped the vehicle's dynamical data into clusters. These clusters were found to map one-to-one with the vehicle's dynamical states, including clusters that mapped to faults.

Fagogenis et al. (Fagogenis, De Carolis, & Lane, 2016) used a Bayesian model with a hidden switch variable to detect partial loss of AUV thrust. As with Raanan et al., training used the vehicle's dynamical sensors, this time to create models of the dynamics and an estimate of the switch variable.

Raanan et al. and Fagogenis et al. both targeted fault detection based on changes in the vehicle's dynamics. Here we focus on the detection of actuator faults. Specifically, we consider an AUV mass-shifter, the device that moves the vehicle's battery back and forth to change pitch, and create an anomaly detector from its input and output current and velocity under nominal and faulty conditions.

The paper is organized as follows. Section 2 describes the mass-shifter, the data collection setup, the data processing,

and the fault-detection algorithm. Section 3 presents the experimental results and the detector's fault-detection performance. Section 4 discusses the results, and Section 5 presents our conclusions.

2. METHODS

2.1. System Description

The vehicle under consideration is a Tethys-class Long-Range Autonomous Underwater Vehicle (LRAUV). LRAUVs perform unmanned basin-scale oceanographic measurements, and have an operational envelope of 21 days (Bellingham et al., 2010).

LRAUVs have 6 actuators (Figure 1): a thruster, elevator and rudder combination for control of the vertical and horizontal planes, an internal mass-shifter for pitch control, a variable buoyancy system, and a drop-weight for emergency recovery. The vehicle carries a standard suite of navigational sensors: depth, 9-axis inertial measurement unit, 3-axis ground-referenced velocity, and a mast-mounted GPS for geo-referencing while at the surface. Communication with the vehicle is done over satellite through an Iridium modem.

The LRAUV's health management system consists of a set of fault-detection and fault-recovery components spread across the individual subsystems, with a common escalation architecture (Kieft et al., 2011). Each component detects faults based on a certain set of threshold-based conditions, and fault-recovery is attempted if the threshold is crossed n consecutive times. If fault-recovery fails, the vehicle ascends to the surface and communicates with a remote operator who then determines whether to clear the fault or abort the mission. Based on 10,000 mission hours, it was found that 95% of the faults are false-alarms and readily cleared by the operator (Bellingham, 2014).

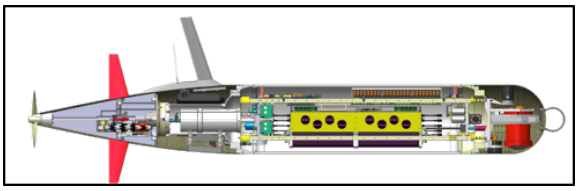

Figure 1. Cutaway of the Long-Range AUV. The mass-shifter (yellow) sits in the vehicle housing.

The focus of this paper is the mass-shifter. The mass-shifter is one of the vehicle's critical sub-systems: although it has a high record of reliability, its failures can cause vehicle loss.

The mass-shifter consists of a DC brushed motor with a planetary gear connected to the battery through a lead screw (Figure 2). The motor's servo-controller runs in constant velocity mode using motor encoder counts for feedback. The known modes of failure of the mass-shifter are current overload, where the tray runs into its travel limit, and loosening of the screw securing the battery to the lead screw, which releases the battery from any constraints.

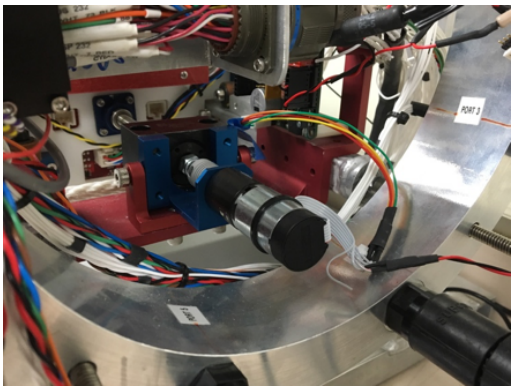

Figure 2. Front view of the LRAUV's mass-shifter: The motor is in the foreground, and a lead screw connects the battery to the motor shaft.

2.2. Data Collection

We instrumented a mass-shifter using high-resolution sensors. We used a National Instruments shunt current sensor inline with the motor (NI-DAQ 9227), a National Instruments voltage sensors across the motor terminals (NI-DAQ 9229), an absolute battery tray position sensor (Tensor Solutions SP1-4 analog string potentiometer, powered by 10V isolated supply, sampled by NI-DAQ 9229), and a pair of vibration accelerometers mounted on the battery tray (PCB-Piezotronics 622B01 and 333B40 ICP 1-axis piezo accelerometers, sampled by IEPE-enabled NI-DAQ 9234). Sensor data was passed through 100 dB anti-aliasing filters at the Nyquist frequency, and sampled at 24 bits. Time-synchronization was maintained by a National Instruments Compact-DAQ 9174 chassis.

We collected data in three cases:

- Nominal operation: the mass-shifter was commanded to move in forward or in reverse between two positions.
- Known faults: Limit fault, a current overload condition, was simulated by starting adjacent to a travel limit, and motion was commanded until the servo-controller detected a current overload and de-energized the actuator. Set-screw fault was simulated by loosening the screws connecting the lead screw to the battery tray.
- Novel faults: new failure modes were created to assess the detector's response to novelty. We simulated two modes: Sensor fault, where the position sensor was disconnected, and Config fault, where an inadvertent configuration file parameter change was simulated, specifically setting the commanded velocity to 50% and 150% of nominal.

Data was collected in the following order: nominal, limit

fault, nominal, setscrew fault, nominal, sensor fault, nominal, speed configuration fault. Each set was repeated for 5 pitch values. The entire sequence was repeated twice. A total of 5000 seconds of steady-state data was collected, equivalent to 500 hours of at-sea operation.

2.3. Data Processing

Data was processed as follows:

- Current was segmented into startup transient, steady-state, and stop transient. Transient data was discarded, and steady-state current was segmented into non-overlapping 100ms sections. The mean and standard deviation were computed over each segment. The effect of pitch was removed:

$$I = I_{\text{not corrected}} - K \cdot \theta \tag{1}$$

  where the coefficient K = 1.07mA/deg was calculated by linear regression of the nominal current versus pitch. A total of 50,000 data points were generated.

- Position was processed similarly to current except for pitch de-trending.

- Mean velocity and standard deviation were calculated using centered differencing of the mean position. Analysis of error vs sampling period indicated that the measurement error was 0.5μm/s at 10Hz.

- Voltage was processed identically to current. Since motor voltage is linearly dependent on current and velocity, it was not used by the detector.

- Vibration: Compared with the current spectrogram, the vibration spectral lines were diffuse and less numerous, and were not retained.

2.4. Anomaly Detector

There is a reasonable expectation that steady-state data should be clustered and representable by a multi-variate Gaussian model. A multi-variate Gaussian model represents the data with a Gaussian probability density:

$$f(X) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(X-\bar{X})^{+}\Sigma^{-1}(X-\bar{X})\right) \tag{2}$$

where the mean $\bar{X}$ and covariance Σ of the distribution are computed from a training set containing N nominal data points:

$$\begin{aligned} \bar{X} &= \frac{1}{N}\sum_{i=1}^{N} X_i \\ \Sigma &= \frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})(X_i-\bar{X})^{+}. \end{aligned} \tag{3}$$

Testing for normality can be done a number of ways (Mardia, 1980). We used an approach where the probability of occurrence outside regions of increasing Mahalanobis distance is compared with the normal distribution.

The decision whether a measurement is nominal or faulty is based on whether it falls inside or outside the decision boundary:

$$f(X) = \text{threshold}. \tag{4}$$

The decision boundary of a Gaussian detector is an ellipsoid.

To determine the threshold, we used a mixture of nominal and faulty data the development set and created a grid of threshold values. For each value, we computed the probability of detection P_d (probability that a fault is detected as a fault i.e. that it falls outside the decision boundary), the probability of false-alarm P_{fa} (probability that nominal data falls inside the decision boundary), and a score defined by:

$$F = \frac{2}{\frac{1}{P_d} + \frac{1}{1-P_{fa}}}, \tag{5}$$

and selected the threshold that optimized the score.

Performance was measured by computing P_{fa} and P_d on a third set, the test set, consisting of a mixture of nominal and faulty data.

Performance was also measured on a fourth set, the novelty set, consisting of faults not seen by the detector during training. Because all novelty data is faulty, the relevant performance metric is P_d.

To create the training, development, and test sets, data was randomized and split according to:

- Training set: 60% of the nominal data, used to train the detector.
- Development set: 20% of the nominal data and 50% of the faulty data, used to optimize parameters.
- Test set: remaining 20% of the nominal data and remaining 50% of the faulty data, used to measure performance on new data.

Randomization was done using independent random permutations of the data. This procedure is appropriate because, ex-

cept for the data during startup, consecutive samples showed no correlations.The novelty set was not randomized.

A common method to increasing the mean time-to-false-alarm (TTFA) is to use n consecutive detections before calling an alarm:

$$\text{TTFA} = \frac{1}{P_{fa}(n)} \tag{6}$$

where $P_{fa}(n)$ is the probability of encountering n consecutive false alarms. To measure $P_{fa}(n)$, we went over each file, counted the number of points part of a block of n or more false-alarms, and divided this by the total number of points.

An upper bound on the TTFA is provided by uncorrelated data. In this case, the probability of occurrence of n consecutive false-alarms is:

$$P_{fa}(n) = (P_{fa})^n \qquad \text{uncorrelated data} \tag{7}$$

i.e. the TTFA increases exponentially fast with n.

We used a second anomaly detector to assess the performance of the Gaussian detector. We used a 1-class Support Vector Machine (Schölkopf, Platt, Shawe-Taylor, Smola, & Williamson, 2001), which maps data into a high-dimensional feature space via a kernel, and then iteratively finds the surface that maximizes the margin between nominal and faulty data. We used a Gaussian kernel because it produces more robust classification models than other functions like polynomial or sigmoidal kernels (Bounsiar & Madden, 2014).

The 1-class SVM has two hyper-parameters. To select their value we performed a grid search where for each parameter pair we computed the decision boundary using the training set, then computed the score described in Equation (5) using the development set, and finally kept the parameter pair that optimized the score.

3. RESULTS

Figure 3 shows the actuator current vs time before processing. Current started with a 500mA / 250ms startup transient, followed by steady-state at 20mA, and ended with a 500mA / 250ms decay transient. Velocity (Figure 4) exhibited an undershoot during the initial rise, followed by steady-state at 0.7mm/s steady-state. The velocity overshoot was observed in the string-pot velocity but not in the motor encoder's, indicating differential motion between motor shaft and mass at startup possibly due to a temporary deformation of the mass-shifter wheels. The current spectrogram (Figure 5) had showed a series of spectral lines. Lines were observed at the fundamental, 1st harmonic, and 7th harmonic of the motor revolution rate (67rps). Multiple lines were also observed at non-integer multiples, believed to originate with the motor gearhead. Spectral information, albeit useful for diagnostic, was not used for anomaly detection.

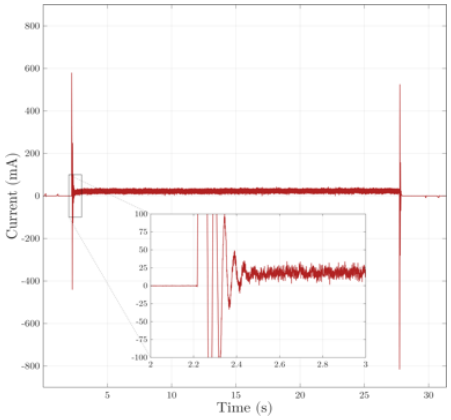

Figure 3. Current versus time. The insert shows details of the startup transient.

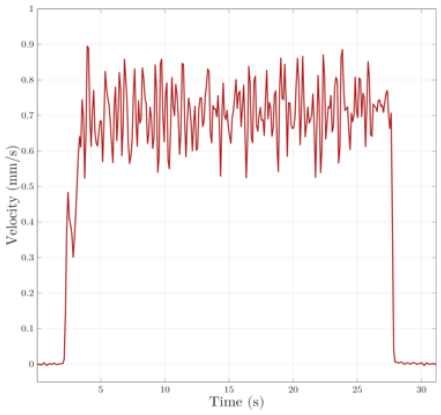

Figure 4. Velocity versus time.

Figure 6 shows the current vs velocity under nominal conditions at pitch = 0. The two clusters correspond to forward motion (upper quadrant) and reverse (lower quadrant). The first second of motion is shown with orange stars: as explained in the previous paragraph, the tail is due to string-pot velocity overshoot.

Testing for normality of the data was done by comparing the fraction of nominal points outside the Mahalanobis distance

$$D(X) = \sqrt{(X - \bar{X})^{+}\Sigma^{-1}(X - \bar{X})} \tag{8}$$

with the prediction for a 2D Gaussian

$$P(D(X) > A) = \exp\left(-\frac{1}{2}A^2\right). \tag{9}$$

The two curves diverged at A=2.5, indicating that 95% of the points were well-represented by a normal distribution. The

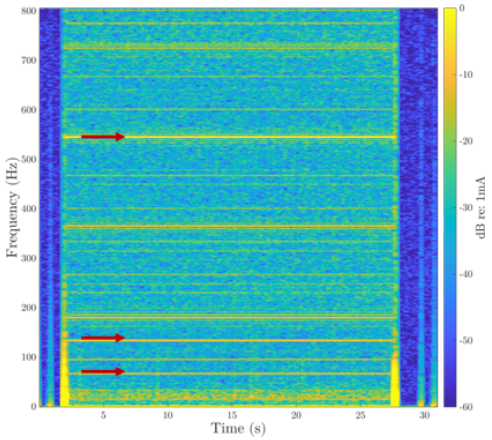

Figure 5. Current spectrogram. Red arrows correspond to harmonics of the motor rate.

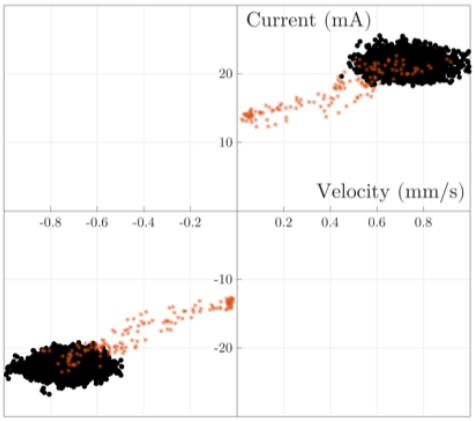

Figure 6. Current vs velocity under nominal conditions at pitch=0. Data corresponding to the first second of motion is shown with orange stars. Top right: forward. Bottom left: reverse.

remaining 5% corresponded mostly to string-pot undershoot during startup.

Figure 7 shows the complete set of nominal and faulty data under all pitch conditions in the forward direction, and the decision boundary the reverse direction data is a mirror image and is not shown. We observed the following:

- The nominal data was well-clustered. As explained above, it is well modeled by a 2D Gaussian except for the 5% of the data associated with string-pot startup undershoots.
- The limit fault data spread from the nominal cluster toward high current/low velocity. This is as expected because the actuator is pressing against a hard mechanical stop. Most of the data fell well outside the decision boundary, indicating a high probability of detection.
- Careful examination of the set-screw fault data indicated that separation did not occur when the actuator pushed against gravity. Since in this case the fault was not expressed, this data was discarded. The remaining data fell around the velocity = 0 line, well outside the decision boundary.
- The speed fault data had two clusters, one at 50% of nominal velocity, one at 150% – as expected. Most of the data fell outside the decision boundary.
- The sensor fault data had a single cluster at zero velocity, well outside the decision boundary.

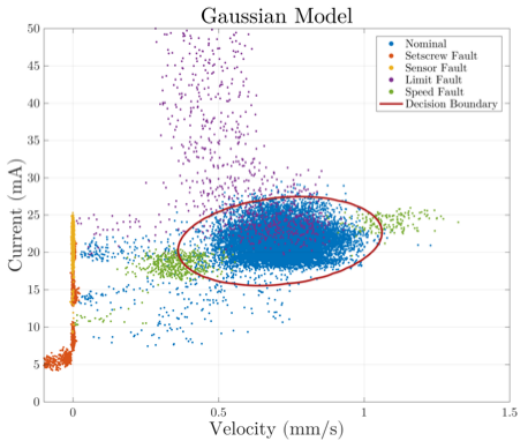

Figure 7. Distributions of features under nominal and faulty conditions, and Gaussian model decision boundary. Color coding is detailed in the legend box.

For comparison, Figure 8 shows the 1-class SVM decision boundary. The performance is comparable, except that the SVM gives more importance to the 5% of points in the tail. As a result, the SVM has a slightly smaller rate of false-alarm but a smaller probability of detection.

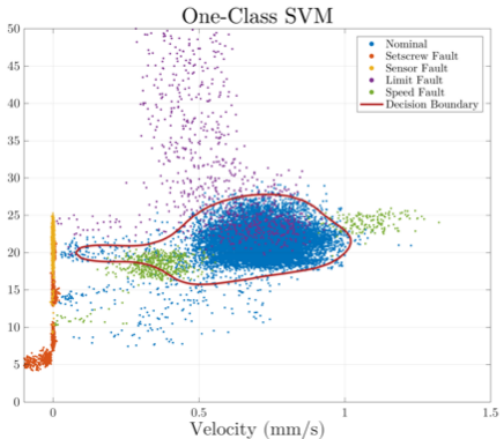

Figure 8. Decision boundary for 1-class SVM.

Figure 9 shows the ROC curve and the optimal threshold location for the Gaussian model and comparison with the 1-class

SVM. The optimal threshold was 6x10-3, corresponding to a Mahalanobis distance of 3.

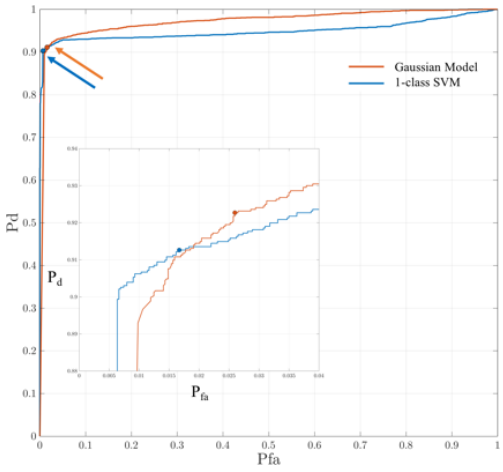

Figure 9. ROC curve for the Gaussian detector (red), and comparison with a 1-class SVM (blue). Optimal performance occurred near Pd = 0.9. (red and blue circles). The insert shows a zoom around the optimal section.

Table I shows that the single-point P_{fa} and P_d of the two detectors was comparable: P_{fa} 2% and P_d 90% on the test set, and P_d 90% on the novelty set.

Table 1. Detection algorithm performance.

Set	Gaussian		1-class SVM	
	Pd%	Pfa%	Pd%	Pfa%
Dev Set	92	2.6	91	1.7
Test Set	91	2.2	90	1.8
Novelty Set	92	-	86	-

We examined the effect of additional features and found it to be negligible. For an electric motor, voltage is linearly related to current and velocity i.e. it adds little information. We found that the current and velocity standard deviations were well-clustered during nominal operation, and that they changed on certain faults, however because the mean current and velocity already provided good separation between nominal and faulty data, the effect on P_d and P_{fa} was negligible.

Figure 10 shows the mean number of samples required to encounter n consecutive false-alarms versus n. The Gaussian model and 1-class SVM performed similarly, but neither achieved the growth predicted for uncorrelated data (Equation 7).

4. DISCUSSION

4.1. Summary

We developed an anomaly detector for a mass-shifter using its current (input) and velocity (output). To do so we collected current-velocity data under nominal, faulty, and novel

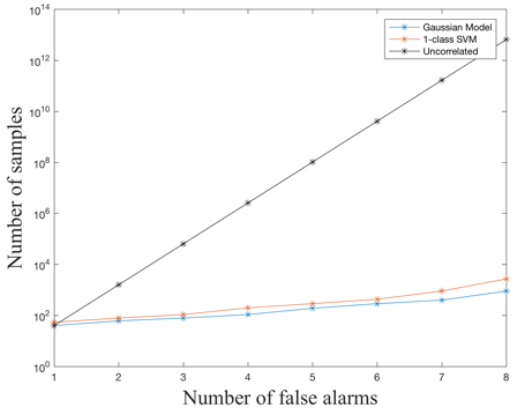

Figure 10. Mean number of samples versus detection latency for the Gaussian model (blue) and the 1-class SVM (red). The uncorrelated case is shown in black.

conditions, and developed a Gaussian anomaly detector. The detector achieved a 90% probability of fault detection and a 2% false-alarm rate. We compared this to a 1-class SVM and found comparable performance. We measured the mean time-to-false-alarm versus number of consecutive false alarms and found that the improvement fell short of the uncorrelated case.

4.2. Performance

The detector's performance is surprisingly good considering that the feature space is only two-dimensional. The fundamental reason for this is that the mass-shifter operates in a highly structured environment with a single set point: it has 1 degree of freedom, it runs closed-loop control on one of the two features, and the input is measured. Additionally, fluctuations due to friction, gear tooth engagement torque, and small shape variations along the screw are small. Under those conditions, even a detector with a simple decision boundary like a Gaussian detector is expected to capture most of the nominal data with a low P_{fa}.

4.3. Temporal Correlation

We found that the mean time-to-false-alarm only increased 20-fold as the criterion changed from 1 to 8 consecutive false-alarms. The expected change assuming uncorrelated data is 1011, indicating that the false-alarms were temporally correlated.

As discussed in the results section, a velocity undershoot was observed during the startup transient. This overshoot lasted anywhere from 0.5 to 1.5 seconds depending on pitch and direction of motion, and sometimes was absent. During the overshoot, the feature vector moved outside the decision boundary, producing a sequence of 5 to 15 consecutive false alarms.

Because the longest observed sequence was 15, one could argue that anything over n=15 should be sufficient to eliminate false-alarms. The difficulty with this argument is that in actual operation the duration of motion ranges anywhere from 1 to 30s, i.e. the strategy is not viable over the full operational range.

Although the performance of the detector is encouraging, improvements to its short-duration performance are desirable. Such improvements could be achieved using time-series forecasting algorithms such as auto-regressive models.

4.4. Application to Other Actuators

We are in the process of extending this approach to the LRAUV's other actuators. The methodology is expected to work unchanged on the thruster and on the variable buoyancy system, since both are driven by DC motors and operate in highly structured environments around a small number of set points.

The methodology however will require adaption in order to be applicable to the rudder and elevator. Like the mass-shifter, these actuators run off DC motors but because the set point is continuously updated, their steady-state current and velocity cannot be used. Alternative features capable of capturing system dynamics with a small number of parameters will have to be devised instead.

5. CONCLUSION

We developed an unsupervised anomaly detector for a mass-shifter using its input (current) and output (velocity). The detector achieved 90% probability of fault detection and 2% false-alarm rate. Despite the good single-point performance, we found that the n-point false-alarm rate scaled poorly with n due to correlation during actuator startup. Improvements to the short-duration performance could be achieved with time-series forecasting.

ACKNOWLEDGMENT

This work was done with financial support from the Packard Foundation.

REFERENCES

Bellingham, J. G. (2014). *Fault detection, diagnosis, and mitigation for long-duration auv missions with minimal human intervention.* (ONR Autonomy Workshop, Sept 25, Arlington VA.)

Bellingham, J. G., Zhang, Y., Kerwin, J. E., Erikson, J., Hobson, B., Kieft, B., & Banka, A. (2010). Efficient propulsion for the tethys long-range autonomous underwater vehicle. In *Proceedings of ieee/oes autonomous underwater vehicles conference.*

Bounsiar, A., & Madden, M. G. (2014). Kernels for one-class support vector machines. In *Proceedings of ieee international conference on information science and applications.*

Brito, M. P. (2015). *Reliability case notes no 9: Autosub long range risk assessment report* (Tech. Rep. No. 51). NOCR.

Brito, M. P., Smeed, D., & Griffiths, G. (2014). Underwater glider reliability and implications for survey design. *Journal of Atmospheric and Oceanic Technology*, *31*, 2858.

Fagogenis, G., De Carolis, V., & Lane, D. M. (2016). Online fault detection and model adaptations for underwater vehicles in the case of thruster failures. In *Ieee international conference on robotics and automation.*

Griffiths, G., Millard, N. W., McPhail, S. D., Stevenson, P., & Challenor, P. G. (2003). On the reliability of the autosub autonomous underwater vehicle. *International Journal of the Society for Underwater Technology*, *25*, 175.

Kieft, B., Bellingham, J. G., Godin, M. A., Hobson, B. Hoover, T., McEwen, R. S., & Mellinger, E. C. (2011). Fault detection and failure prevention on the tethys long-range autonomous underwater vehicle. In *Proceedings of unmanned untethered submersible technology conference, august 21-24, durham nh.*

Mardia, K. V. (1980). 9 tests of univariate and multivariate normality. *Handbook of Statistics*, *1*, 279.

Raanan, B., Bellingham, J. G., Zhang, Y., Kemp, M., Kieft, B., Singh, H., & Girdhar, Y. (2016). Automatic fault diagnosis for autonomous underwater vehicles using online topic models. In *Proceedings of ocean 2016 conference, sept 19-22, monterey ca.*

Raanan, B., Bellingham, J. G., Zhang, Y., Kemp, M., Kieft, B., Singh, H., & Girdhar, Y. (2017). Detection of unanticipated faults for autonomous underwater vehicles using online topic models. *Journal of Field Robotics*.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural computation*, *13*(7), 1443-1471.

BIOGRAPHIES

Dr. Matt Kemp is a Principal Engineer at the Monterey Bay Aquarium Research Institute in Moss Landing CA. He holds a Ph.D. in Physics from the University of North Carolina at Chapel Hill (92). Dr. Kemp served as Director of Concept Development with Bluefin Robotics for 5 years, and as Director of Concept Development with Nekton Research for 7. His research interests are AUV design and vehicle health management. He is a member of IEEE..

Ben Raanan is a Research Specialist at the Monterey Bay Aquarium Research Institute in Moss Landing CA. He received his B.Sc. in Marine Science from the Ruppin Academic Center (Israel) and a M.S. in Physical Oceanography

from Moss Landing Marine Laboratories. His research interests are AUV autonomy, machine learning and artificial intelligence.