

Unsupervised Anomaly Detection for Self-flying Delivery Drones

Vikas Sindhvani¹, Hakim Sidahmed¹, Krzysztof Choromanski¹, Brandon Jones²

Abstract— We propose a novel anomaly detection framework for a fleet of hybrid aerial vehicles executing high-speed package pickup and delivery missions. The detection is based on machine learning models of normal flight profiles, trained on millions of flight log measurements of control inputs and sensor readings. We develop a new scalable algorithm for robust regression which can simultaneously fit predictive flight dynamics models while identifying and discarding abnormal flight missions from the training set. The resulting unsupervised estimator has a very high breakdown point and can withstand massive contamination of training data to uncover what normal flight patterns look like, without requiring any form of prior knowledge of aircraft aerodynamics or manual labeling of anomalies upfront. Across many different anomaly types, spanning simple 3-sigma statistical thresholds to turbulence and other equipment anomalies, our models achieve high detection rates across the board. Our method consistently outperforms alternative robust detection methods on synthetic benchmark problems. To the best of our knowledge, dynamics modeling of hybrid delivery drones for anomaly detection at the scale of 100 million measurements from 5000 real flight missions in variable flight conditions is unprecedented.

I. INTRODUCTION

As aerial robots [1], [2] become increasingly capable of complex navigation, perceptual reasoning and ability to learn from experience, it is expected that a large number of delivery missions will soon be executed by small air-vehicles taking off autonomously, flying far beyond line of sight over densely populated areas, hovering inside a residential zone within touching distance of humans to deliver the package, and returning to their “home” upon mission completion. Ensuring as high degree of operational reliability and safety as passenger airplanes is critical for delivery drones to achieve economies of scale.

While simple statistical thresholds and logical rules can be hand-designed to trigger on frequently occurring problematic events (e.g., battery too low, or control surface non-functional), they cannot exhaustively cover all potential future failure modes which are unknown a priori, particularly as the fleet grows in mission complexity and vehicle types. With this motivation, we develop an anomaly detection system based on a machine-learning model that is continuously trained on thousands of flight logs. When this model reports large predictive errors for a new flight, the vehicle can be flagged for manual inspection and possibly grounded for safety until the issue is resolved. Importantly, the system is designed to *discover normality* and does not require upfront labeling of normal and anomalous missions; indeed, sifting

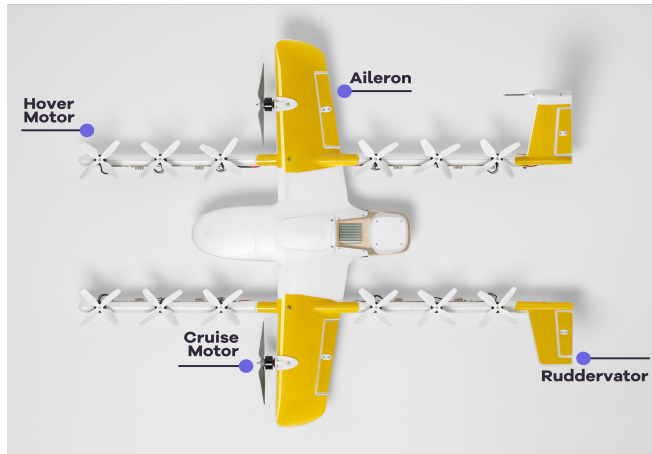


Fig. 1. Hybrid Aerial Vehicle for High-speed Package Pickup and Delivery

through thousands of flight logs comprising of dozens of time series looking for subtle abnormalities stretches the limits of what is manually feasible.

As in prior work [3], [4], [5], [6] on fault detection and fleet monitoring of aircrafts [7], [8], [9], our framework relies on learning a predictive model of flight dynamics. The linear and angular acceleration of an aircraft depends on the aerodynamic forces it is subject to, which are a function of the vehicle state, control commands, dynamic pressure and other flight condition variables. As [3] show, simple linear or quadratic models trained on historical flight logs show impressive predictive power. The norm of the predictive residuals at a given time for a given flight, or the mean residual over an entire flight can be used as thresholds for anomaly detection. However, in contrast to prior work that focused on large fixed wing passenger aircrafts and cruising performance only, we are interested in monitoring much smaller delivery drones [10] across an entire flight mission that includes takeoff, package delivery, and landing. Furthermore, to span a full range of conditions from stable hovering to energy efficient cruising, we work with a hybrid small air-vehicle schematically shown above in Figure 1. An array of 12 vertically mounted electric motors provides thrust for hovering flight. Two forward thrust motors, two ailerons, and two ruddervators are used primarily for cruise flight.

This hybrid configuration makes the task of building an accurate model of the system more challenging as the aerodynamic interactions are more complex than on larger fixed-wing aircraft (e.g. rotor cross-flow, flow around small structures). As an alternative to pushing the boundary of computational fluid dynamics tools, or performing complex

¹Robotics at Google, New York Email: {sindhvani, hsidahmed, kchoro}@google.com

²Wing Aviation Email: brnjones@wing.com

and expensive measurement campaigns using wind tunnels, learning models from raw flight data turns out to be surprisingly effective.

Trained on among the largest scale real-world delivery drone data reported to date, our detectors successfully flag missions with disabled actuators, off-nominal hardware conditions, turbulence and other anomalous events. Our approach is based on a combination of non-parametric dynamics modeling and a novel algorithm for robust and scalable least trimmed squares estimation, which may be of independent interest.

II. DYNAMICS LEARNING AND ANOMALY DETECTION

In this section, we introduce notation and formulate the problem abstractly. Consider a robot interacting with its environment according to an unknown continuous-time nonlinear dynamical system,

$$\dot{x}(t) = f(x(t), u(t))$$

where states $x(t) \in \mathbb{R}^n$ and controls $u(t) \in \mathbb{R}^m$. Assume that a fleet of such robots collectively and continuously execute missions generating trajectory logs of the form,

$$\tau_i = \{(x_i(t), u_i(t), \dot{x}_i(t))\}_{t=0}^{T_i}$$

where i indexes the mission.

From N mission logs, one may naturally hope to learn f over a suitable family of function approximators \mathcal{F} by solving a least squares problem,

$$f^* = \arg \min_{f \in \mathcal{F}} \sum_{i=1}^N r(\tau_i, f) \quad (1)$$

where r denotes the predictive residual,

$$r(\tau, f) = \frac{1}{T} \sum_{t=1}^T \|\dot{x}(t) - f(x(t), u(t))\|_2^2$$

While this is reminiscent of model-based Reinforcement Learning, our interest in this paper is not to learn controllers, but rather to turn the dynamics estimate f^* into a detector that can flag mission abnormalities. For any trajectory τ generated by a new mission, the per time-step residual norm $\|\dot{x}(t) - f^*(x(t), u(t))\|_2^2$ is a measure of “instantaneous unexpectedness” and the mean residual across time, $r(\tau, f^*)$, defines an anomaly score for that mission.

Chu et. al. [3] adopt this approach for predicting linear and angular acceleration of the aircraft. By using linear and quadratic functions, a single pass over the mission logs suffices for least squares estimation.

In practice, such an approach to anomaly detection may become fragile in the face of the quality of real world data. When the set of training missions is contaminated with operational failures or carry subtle signatures of future catastrophes (e.g., sensor degradation), the detector may extract a misleading characterization of normal behavior. Unlike model-based RL settings where all collected trajectories may be useful for learning the unknown dynamics, for anomaly detection the learning process has to simultaneously filter

out missions for such abnormalities while fitting a model to the data that remains. In the absence of such a filtering mechanism, it is well known that ordinary least squares estimators and associated anomaly detectors may degrade in quality due to the presence of highly abnormal missions in the training set.

III. ROBUST DYNAMICS LEARNING

A measure of robustness of an estimator is the *finite-sample breakdown point* [11] which in our context is the fraction of mission trajectories that may be arbitrarily corrupted so as to cause the parameters of the estimator to blowup (i.e., become infinite). For least squares estimators or even least absolute deviations (l_1) regressors, the finite sample breakdown point is $\frac{1}{N}$ making them fragile in the presence of heavy outliers in the training set. A more robust alternative is trimmed estimators. For any f , denote the order statistics of the residuals as,

$$r(\tau_{[1]}, f) \leq r(\tau_{[2]}, f) \leq \dots \leq r(\tau_{[N]}, f)$$

Then we define the *trimmed estimator* [12], [11], [13] as the sum of the smallest k residuals,

$$f^* = \arg \min_{f \in \mathcal{F}} \sum_{i=1}^k r(\tau_{[i]}, f) \quad (2)$$

The breakdown point of such an estimator is $\frac{N-k+1}{N}$ where k is the number of missions that should not be trimmed. In practice, k is unknown and is treated as a hyper-parameter. By making k small enough, the breakdown point can even be made larger than 50%.

The price of strong robustness is computational complexity of least trimmed squares estimation [14]: for an exact solution, the complexity scales as $O(N^{d+1})$ for $d \geq 3$ dimensional regression problems. The optimization task is both non-smooth and non-convex. Due to its combinatorial flavor, it is not amenable to standard gradient techniques or least squares solvers even for linear models. Thus, the development of practical approximate algorithms [15], [16] is of significant interest. We now develop a novel algorithm for robust learning based on smoothing the trimmed squares loss. The algorithm is inspired by Nesterov’s smoothing procedure for minimizing non-smooth objective functions [17], and is also closely related to Deterministic Annealing [18], [19], [20] methods for combinatorial optimization.

A. Smoothing the Trimmed Loss

Consider the function that maps a vector $r \in \mathbb{R}^N$ to the sum of its k smallest elements,

$$h_k(r) = \sum_{i=1}^k r_{[i]} \quad \text{where } r_{[1]} \leq r_{[2]} \leq \dots \leq r_{[N]}$$

This function admits a smoothing [21], [22] defined as follows,

$$h_k^T(r) = \min_{\alpha \in \mathbb{R}^N} \alpha^T r + T \sum_{i=1}^N H(\alpha_i) \quad (3)$$

$$\text{s.t:} \quad \sum_{i=1}^N \alpha_i = k, \quad 0 \leq \alpha_i \leq 1 \quad (4)$$

$$\text{where} \quad H(u) = u \log(u) + (1-u) \log(1-u) \quad (5)$$

Above, T is a smoothing parameter also referred to as the ‘‘temperature’’ in the annealing literature. Intuitively, if α_i tends to zero, the corresponding mission is considered too anomalous for training and is trimmed away. The α 's may also be interpreted as probability distributions over binary indicator variables encoding whether or not to trim a mission. As such, when T is high, the smoothed objective is dominated by the entropy of the α 's and tend to approach the uniform distribution $\forall i: \alpha_i = \frac{k}{N}$. As $T \rightarrow 0$, the weights harden towards binary values. This strategy of starting with a highly smoothed proxy to a non-convex non-smooth function and gradually increasing the degree of convexity is the central idea of homotopy [23], continuation [24] and graduated non-convexity [25], [26] methods for global optimization. In the ideal case, the highly smoothed function is close to being convex allowing the global minimum to be found efficiently. As smoothing is reduced, one hopes that following the continuous path of the minimizer would lead to the global minimum. Figure 2 shows how spurious local minima can be eliminated due to smoothing, making the optimization task much easier.

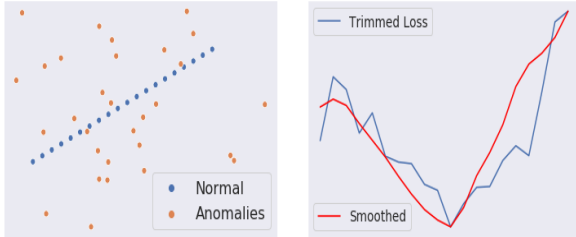


Fig. 2. Left: Normal data follows $y = mx$ where the slope is $m = 1$. Right: trimmed loss and its smoothing ($T = 0.025$) as a function of m .

In particular, the smoothing discussed above has the following properties [21], [22]:

- h_k^T is a concave function.
- h_k^T is continuously differentiable.
- $h_k^T(r) - TR \leq h_k(r) \leq h_k^T(r)$ holds for some fixed constant R .

B. Optimizing the Smoothed Trimmed Loss

With this smoothing of the trimmed loss, for a fixed number of k missions to retain, we consider the following optimization problem,

$$f^* = \arg \min_{f \in \mathcal{F}} h_k^T(r(f)), \quad r(f) = [r(\tau_1, f) \dots r(\tau_N, f)]$$

Equivalently,

$$f^* = \arg \min_{f \in \mathcal{F}, \alpha \in \mathbb{R}^N} \sum_{i=1}^N \alpha_i r(\tau_i, f) + TH(\alpha_i) \quad (6)$$

$$\text{s.t:} \quad \sum_{i=1}^N \alpha_i = k, \quad 0 \leq \alpha_i \leq 1 \quad (7)$$

We now describe a procedure that alternates between a fitting phase and a trimming phase up to convergence. Both these phases are fast, efficient and easily scale to thousands of missions and millions of measurements. We initialize the optimization with $\alpha = \frac{k}{N}$ which corresponds to the non-robust least squares estimator and the limit of $T \rightarrow \infty$.

1) *Fitting Phase*: We consider linear combinations of fixed nonlinear basis functions,

$$f(x, u) = W\phi(x, u)$$

where $\phi: \mathbb{R}^{n+m} \mapsto \mathbb{R}^d$ is a nonlinear feature map and W is a $n \times d$ parameter matrix.

For fixed α 's, optimizing W is a weighted least squares problem which admits a fast single pass solution,

$$W = [A + \lambda I_d]^{-1} B \quad \text{where,} \quad (8)$$

$$A = \sum_{i=1}^N \alpha_i \sum_t \phi(x_i(t), u_i(t)) \phi(x_i(t), u_i(t))^T \quad (9)$$

$$B = \sum_{i=1}^N \alpha_i \sum_t \phi(x_i(t), u_i(t)) \dot{x}_i(t)^T \quad (10)$$

2) *Trimming Phase*: For fixed W , we compute the vector of N residuals given by $r_i = r(\tau_i, W)$. The α optimization takes the form [17],

$$\alpha_i = \frac{1}{1 + \exp\left(\frac{r_i - \nu}{T}\right)} \quad (11)$$

where the scalar ν satisfies the nonlinear equation,

$$\psi(\nu) = \sum_{i=1}^N \frac{1}{1 + \exp\left(\frac{r_i - \nu}{T}\right)} - k = 0 \quad (12)$$

The root of this equation can be easily solved e.g., via the bisection method noting that $\psi(a) < 0$ for $a = \min_i r_i - T \log \frac{N-k}{N}$ and $\psi(b) > 0$ for $b = \max_i r_i - T \log \frac{N-k}{N}$ provides an initial bracketing of the root.

C. Nonlinear Dynamics via Random Fourier Features

We experimented with both linear models as well as nonlinear random basis functions [27], [28] of the form,

$$\phi(x, u) = \sqrt{\frac{2}{d}} \cos(\sigma^{-1} Gx + \sigma^{-1} Hu + b) \quad (13)$$

$$\text{where} \quad G_{ij}, H_{ij} \sim \mathcal{N}(0, 1), b \sim \mathcal{U}(0, 2\pi)$$

$$\text{and} \quad G \in \mathbb{R}^{d \times n}, H \in \mathbb{R}^{d \times m} \quad (14)$$

Here, the feature map dimensionality d controls the capacity of the dynamics model. In particular, as $d \rightarrow \infty$, inner products in the random feature space approximate the Gaussian

Kernel [29],

$$\phi(x, u)^T \phi(\bar{x}, \bar{u}) \approx e^{-\frac{\|x - \bar{x}\|_2^2 + \|u - \bar{u}\|_2^2}{2\sigma^2}}$$

The implication of this approximation is that each component of the learnt dynamics function is a linear combination of similarity to training mission measurements, in the following sense.

$$f_j(x, u) = w_j^T \phi(x, u) \approx \sum_{i=1}^N \sum_t \beta_{j,i,t} e^{-\frac{\|x - x_i(t)\|_2^2 + \|u - u_i(t)\|_2^2}{2\sigma^2}}$$

for some coefficients $\beta_{j,i,t}$ and where w_j^T refers to the j -th row of W . However, the random feature method scales linearly in the number of measurements, as opposed to cubically for β when working with exact kernels. At the price of losing this linear training complexity and globally optimal solution, one may also embrace deep networks for this application to parameterize the dynamics model.

D. Comparison on Synthetic Missions

We generate synthetic 8-dimensional input and 3-dimensional output time series following a linear model as follows. The output time series for normal missions carry moderate Gaussian noise, but anomalous missions are heavily corrupted by non-Gaussian noise sampled uniformly from the interval $[0, 10]$. 200 training and 200 test missions, each with 100 time steps were generated with 50% anomalies following the procedure below. The anomaly labels were discarded for training and only used for evaluation.

$$\begin{aligned} x(t) \in \mathbb{R}^8 &= \cos(g * t + b), \quad g, b \in \mathbb{R}^8, g_i, b_i \sim \mathcal{N}(0, 1) \\ y(t) \in \mathbb{R}^3 &= Wx(t) + \epsilon(t), \quad W \in \mathbb{R}^{3 \times 8} \\ \epsilon(t) \in \mathbb{R}^3 &:= \begin{cases} \epsilon_i \sim \mathcal{N}(0, 1) & \text{for normal} \\ \epsilon_i \sim \text{Unif}(0, 10) & \text{for anomaly} \end{cases} \end{aligned}$$

Examples of output profiles in normal and anomalous missions are shown below in Figure 3.

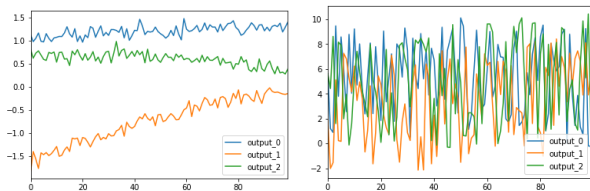


Fig. 3. Normal (left) and Anomalous (right) output profiles.

Figure 4 shows how alternating optimization of the smoothed trimmed loss (with temperature $T=1.0$) leads to monotonic descent in the sum of the smallest $k = 100$ residuals, a consequence of our smoothing formulation. The optimization converges to a set of mission weights (α 's) that clearly trim away nearly all the anomalies present in the training set despite heavy 50% corruption and no explicit anomaly labels provided to the algorithm.

Figure 5 shows a comparison of anomaly detectors based on the proposed trimming approach against pure least squares

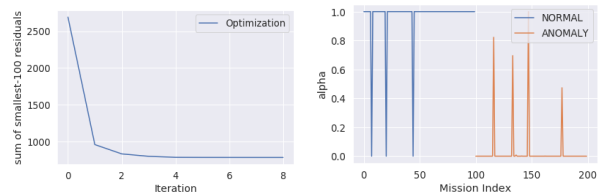


Fig. 4. Objective vs Iterations (left) and Mission Weights (α 's learnt)

models, and several M -estimators proposed in the Robust Statistics [11] literature. Due to its non-robustness, the least squares detector is hurt the most due to corruptions in the training set. By limiting outlier influence, robust loss functions such as l_1 and *huber* show improved performance. Yet, they are outperformed by the proposed trimming approach which gives perfect detection despite massive corruption of the training data.

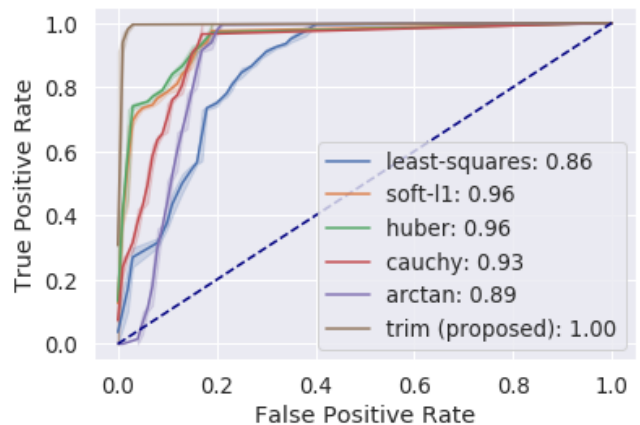


Fig. 5. Smoothed Trimmed (proposed) vs Robust estimators.

IV. ANOMALY DETECTION FOR DELIVERY DRONES

A. Missions

We now report results on data collected from a fleet of delivery drones flying in multiple environments on real delivery flight missions. A typical mission consists of a takeoff, package pickup, a cruise phase to deliver the package, and subsequent landing. To the best of our knowledge, machine learning on real delivery drone data at this scale is unprecedented: 5000 historical missions prior to a cutoff date generating around 80 million measurements are used for training. Trained detectors were tested on 5000 outdoor missions after the cutoff date. By contrast, recent papers have reported results on 20 to 50 test missions in controlled lab environments [30], [31]. Our large-scale flight log data covers multiple vehicle types, package configurations and variable temperature and wind conditions. Additionally, the mission logs are mixed with a variety of research and development flights that include flight envelope expansion, prototype hardware and software, and other experiments designed to stress-test the flight system. Flight missions



Fig. 6. Input Signals for Anomaly Detection

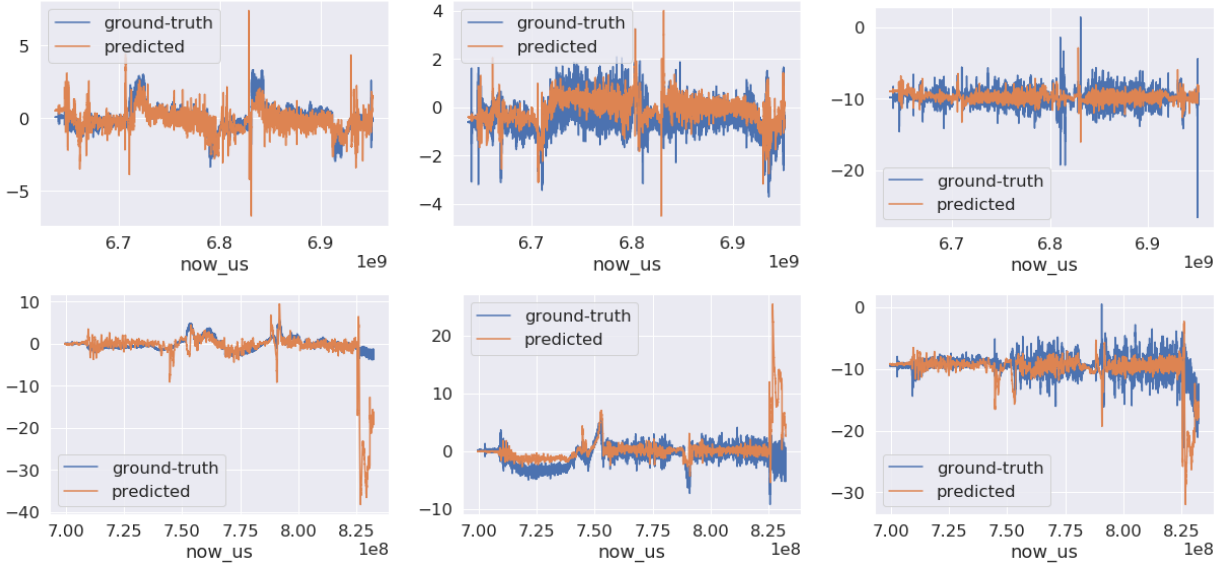


Fig. 7. Detector predictions: x-y-z accelerations for normal (top), and anomalous (bottom) flights with error spikes.

generally last approximately 5 minutes including several kilometers of cruise flight.

B. Signals

Figure 6 shows the input signals used for training models to predict linear and angular acceleration of the vehicle. Each input time series is re-scaled so that values lie in the interval $[-1.0, 1.0]$. Training a nonlinear trimmed model with $d = 100$ random Fourier features on 80 million measurements, including data preprocessing, is completed within 1.15 hours on a single CPU. Predictions on a normal and an anomalous test mission are shown in Figure 7. The large spike towards the end of the anomalous test mission causes the mean residual error to be large, flagging the flight as anomalous.

C. Control

The vehicle’s position, velocity, and attitude estimates from an EKF-based state estimator are compared with commands generated by a high-level mission planning system. The controller generates actuator commands to reduce errors between the state estimate and commands. The controller incorporates a real-time airspeed estimate to properly allocate control between individual hover motors and aerodynamic control surfaces throughout the airspeed envelope.

D. Anomaly Types

We report detection rates for multiple anomaly types:

- **are-basic-stats-exceeded:** Basic statistical measures such as velocity command error, error from commanded

path, pitch, roll, Root Mean Squared pitch and roll error, pitch and roll torque commands are more than 3 standard deviations from the mean computed over the entire training set.

- **has-flight-dynamics-issue:** The particular flight had an issue where the flight dynamics were off-nominal due to various factors such as an intentionally disabled actuator or other off-nominal airframe modifications to test system robustness.
- **is-high-wind:** The prevailing wind speed is greater than 10m/s, which qualitatively indicates elevated levels of turbulence.

Approximately 12% of the test set of 5000 missions has these anomalies.

E. Detection Performance

Figure 8 shows the performance of our nonlinear trimmed detectors ($d = 100, T = 1.0, k = 0.75N$) on a test set of 5000 missions. On all three anomaly types, the area under the True-Positive-Rate vs False-Positive-Rate curve exceeds 0.90. The detector coverage goes beyond simple statistical anomaly measures firing reliably across a multitude of factors such as disabled actuators, otherwise off-nominal hardware conditions and the vehicle experiencing turbulent conditions.

F. Effectiveness of Trimming

Figure 9 shows the smoothed distribution of mission weights learned by the proposed trimming algorithm. The

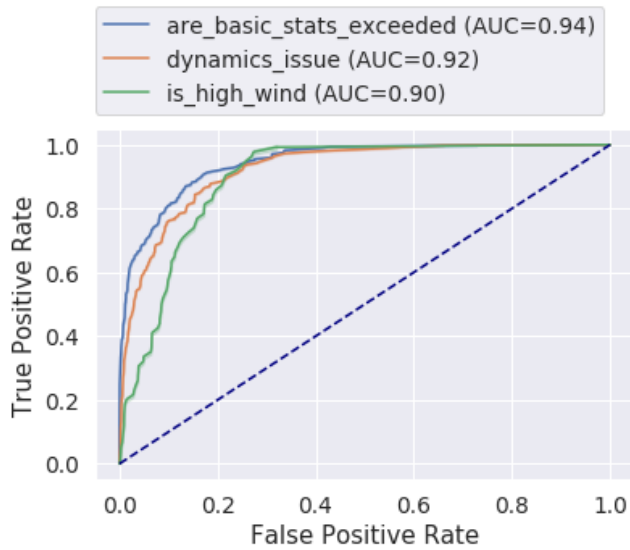


Fig. 8. Anomaly Detection Performance on 5000 Test Missions

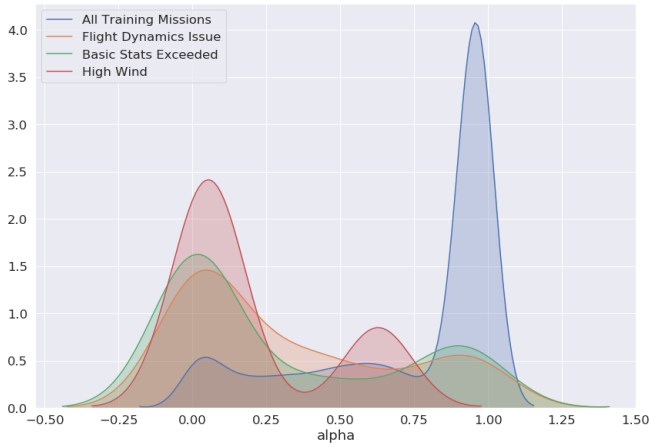


Fig. 9. Distribution of Training Mission Weights

distribution of α 's peaks close to 0 in comparison to the mean over the entire training set which is close to 1.0. This confirms the ability of the proposed method to successfully filter out anomalies from the training set, in order to extract normal flight patterns, without requiring any form of supervision. Without nonlinear modeling and trimming, we observed performance degradation in an analysis across finer anomaly type categories.

G. Case Studies

Two example case studies are presented to illustrate the utility of the anomaly detection algorithm.

1) *Hover Motor Testing*: A test flight was performed with one of the 12 hover motors purposefully disabled. A disabled hover motor is expected to require additional thrust and torque from the remaining hover motors to maintain trimmed flight and therefore represents a different flight dynamics model. The unsupervised anomaly detection algorithm successfully flagged this particular flight as anomalous.

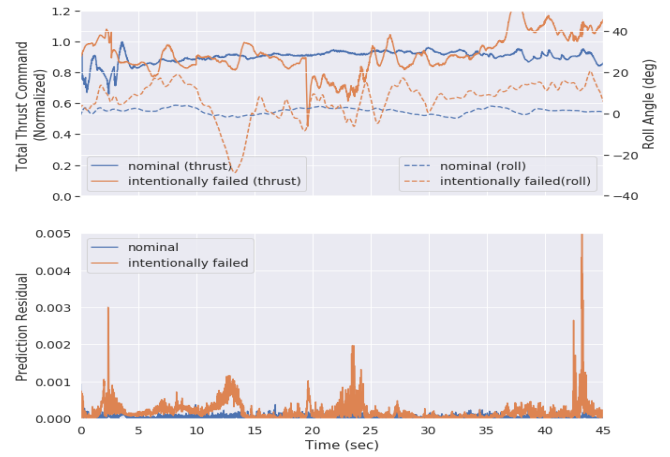


Fig. 10. Nominal Hovering vs Intentionally Failed Hover Motor

Figure 10 shows a comparison between a nominal and the failed hover motor test flight. In this case the prediction residuals are large compared to the nominal flight.

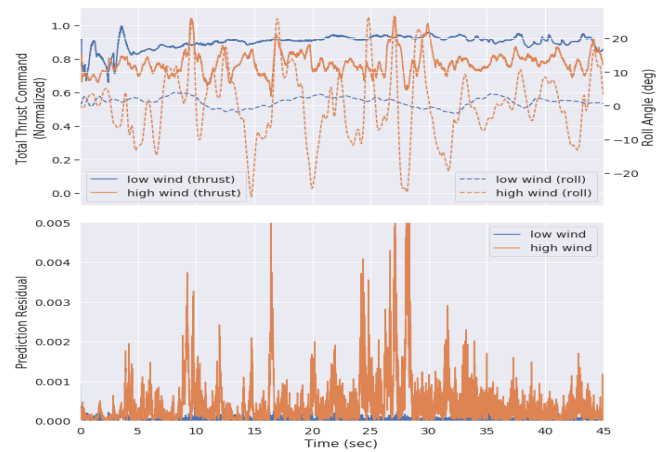


Fig. 11. Low Wind vs High Wind Hovering Dynamics

2) *Impact of Wind Conditions*: We compare two otherwise nominal steady hovering flights differing only in their anomaly score. In this case, the algorithm flagged a flight in high turbulent wind conditions as more anomalous than one in light wind conditions. This is reasonable since high wind represents an unmodeled random disturbance and not intended to be part of the learnt bare airframe flight dynamics model. Figure 11 shows a comparison of the total thrust command and roll angle for each case. The high wind case has a lower overall thrust command due to lift from the wing, but is much more variable as the controller is rejecting the random disturbances. Roll angles in the high wind case are up to 25 deg as the vehicle is maneuvering to maintain its hovering position. Flight data in low-wind conditions is desirable for system identification applications where only the bare-airframe dynamics are desired. This algorithm can be used as a preprocessing step in a large dataset to find data with the least amount of unmodeled disturbances.

REFERENCES

- [1] Kumar and Michael, "Opportunities and challenges with autonomous micro aerial vehicles," in *International Symposium on Robotics Research*, 2011.
- [2] D. Floreano and R. J. Wood, "Science, technology and the future of small autonomous drones," *Nature*, vol. 521, 2015.
- [3] E. Chu, D. Gorinevsky, and S. Boyd, "Detecting aircraft performance anomalies from cruise flight data," in *AIAA Infotech@ Aerospace 2010*, p. 3307.
- [4] D. G. Dimogianopoulos, J. D. Hios, and S. D. Fassois, "Fault detection and isolation in aircraft systems using stochastic nonlinear modelling of flight data dependencies," in *2006 14th Mediterranean Conference on Control and Automation*. IEEE, 2006, pp. 1–6.
- [5] E. Chu, D. Gorinevsky, and S. Boyd, "Scalable statistical monitoring of fleet data," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 13 227–13 232, 2011.
- [6] D. stnevsky, B. Matthews, and R. Martin, "Aircraft anomaly detection using performance models trained on fleet data," in *2012 Conference on intelligent data understanding*. IEEE, 2012, pp. 17–23.
- [7] J. Qi, D. Song, H. Shang, N. Wang, C. Hua, C. Wu, X. Qi, and J. Han, "Search and rescue rotary-wing uav and its application to the lushan ms 7.0 earthquake," *Journal of Field Robotics*, vol. 33, no. 3, pp. 290–321, 2016.
- [8] M. Bragg, T. Basar, W. Perkins, M. Selig, P. Voulgaris, J. Melody, and N. Sarter, "Smart icing systems for aircraft icing safety," in *40th AIAA Aerospace Sciences Meeting & Exhibit*, 2002, p. 813.
- [9] E. Khalastchi and M. Kalech, "On fault detection and diagnosis in robotic systems," *ACM Computing Surveys (CSUR)*, vol. 51, no. 1, p. 9, 2018.
- [10] J. Xu, T. Du, M. Foshey, B. Li, B. Zhu, A. Schulz, and W. Matusik, "Learning to fly: computational controller design for hybrid uavs with reinforcement learning," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, p. 42, 2019.
- [11] R. A. Maronna, R. D. Martin, V. J. Yohai, and M. Salibián-Barrera, *Robust statistics: theory and methods (with R)*. John Wiley & Sons, 2019.
- [12] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. John wiley & sons, 2005, vol. 589.
- [13] P. J. Huber, *Robust statistics*. Springer, 2011.
- [14] D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "On the least trimmed squares estimator," *Algorithmica*, vol. 69, no. 1, pp. 148–183, 2014.
- [15] F. Shen, C. Shen, A. van den Hengel, and Z. Tang, "Approximate least trimmed sum of squares fitting and applications in image analysis," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1836–1847, 2013.
- [16] D. M. Mount, N. S. Netanyahu, C. D. Piatko, A. Y. Wu, and R. Silverman, "A practical approximation algorithm for the lts estimator," *Computational Statistics & Data Analysis*, vol. 99, pp. 148–170, 2016.
- [17] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical programming*, vol. 103, no. 1, pp. 127–152, 2005.
- [18] G. L. Bilbro, W. E. Snyder, and R. C. Mann, "Mean-field approximation minimizes relative entropy," *JOSA A*, vol. 8, no. 2, pp. 290–294, 1991.
- [19] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *science*, vol. 220, no. 4598, pp. 671–680, 1983.
- [20] V. Sindhwani, S. S. Keerthi, and O. Chapelle, "Deterministic annealing for semi-supervised kernel machines," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 841–848.
- [21] Y. Qi, "Smoothing approximations for two classes of convex eigenvalue optimization problems," Ph.D. dissertation, 2005.
- [22] S. Shengyuan, "Smooth convex approximation and its applications," 2004.
- [23] D. M. Dunlavy and D. P. O’Leary, "Homotopy optimization methods for global optimization." Sandia National Laboratories, Tech. Rep., 2005.
- [24] E. L. Allgower and K. Georg, *Numerical continuation methods: an introduction*. Springer Science & Business Media, 2012, vol. 13.
- [25] E. Hazan, K. Y. Levy, and S. Shalev-Shwartz, "On graduated optimization for stochastic non-convex problems," in *International conference on machine learning*, 2016, pp. 1833–1841.
- [26] A. Blake and A. Zisserman, *Visual reconstruction*, 1987.
- [27] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in neural information processing systems*, 2008, pp. 1177–1184.
- [28] H. Avron, V. Sindhwani, J. Yang, and M. W. Mahoney, "Quasi-monte carlo feature maps for shift-invariant kernels," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 4096–4133, 2016.
- [29] B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*, 2001.
- [30] A. Keipour, M. Mousaei, and S. Scherer, "Automatic real-time anomaly detection for autonomous aerial vehicles," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5679–5685.
- [31] —, "Alfa: A dataset for uav fault and anomaly detection," *arXiv preprint arXiv:1907.06268*, 2019.