

ROTOR ROCKSTAR

RED BULL AEROBATICS / **P.50**



AOPATM

EURO BUSHPLANE

STOL MAGNIFIQUE / **P.68**

PILOT

TURBINE EDITION

The Voice of General Aviation
aopa.org/pilot / July 2025 / \$10.95

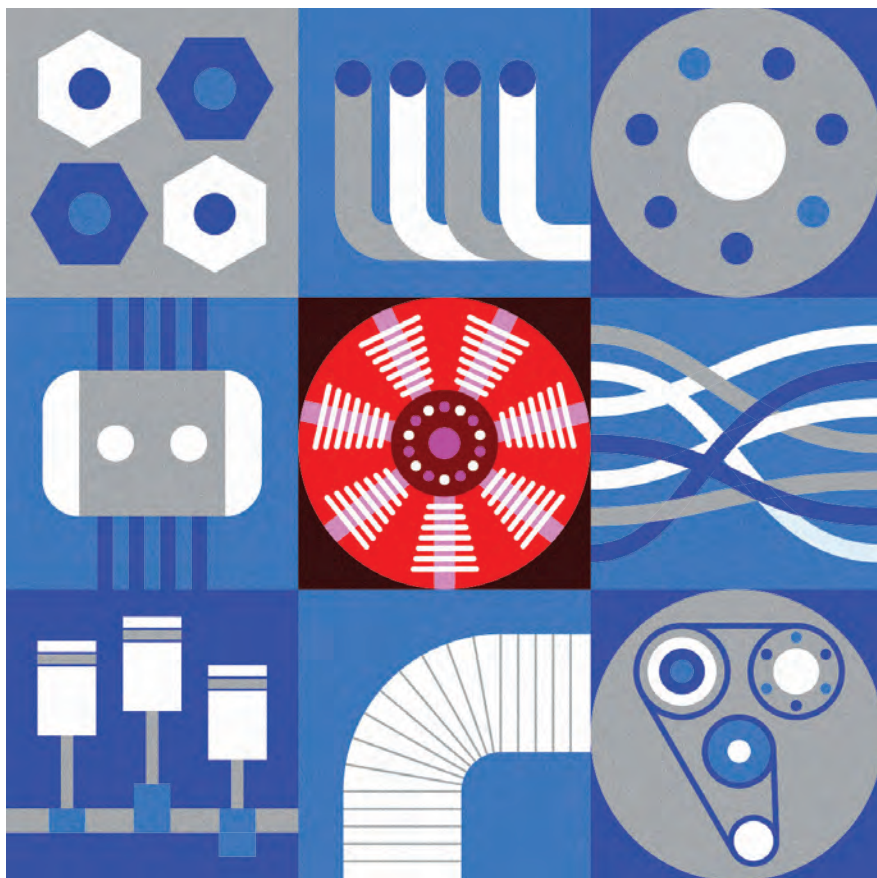


BLACKHAWK TBM 700 UPGRADE

MUSCLE UP

MORE POWER,

BETTER PERFORMANCE / **P.T-10**



SAVVY MAINTENANCE / OPINION

Project GADfly update

Status report on an AI watchdog for digital engine monitor data

BY MIKE BUSCH

I **LAST WROTE** about Project GADfly in the November 2022 issue of *AOPA Pilot* (“When Data Doesn’t Look Right”). This was a first peek at an R&D effort at Savvy Aviation to harness artificial intelligence to analyze the 10,000 flights of digital engine monitor data uploaded to our platform each week. The goal? Identify flights that appear unusual—ones that merit a closer look by our experienced human

data analysts. The name GADfly is short for General Anomaly Detection, with the suffix “fly” an obvious aviation reference.

Savvy has been collecting, archiving, and analyzing digital engine monitor data for piston general aviation airplanes for about 15 years. Although we have some 9 million flights in our database with about 10,000 new ones uploaded every week, only a tiny handful of those flights get reviewed by one

HISTORY TAKES WING

SEE AT OSHKOSH 2025

1929 TRAVEL AIR S6000B

NC9084 IS FOR SALE
IN SEATTLE, WA

INCLUDED IN SALE:

- SPARE WRIGHT WHIRLWIND 975 ENGINE
- BACON 4400 WOOD SKIS AND TAIL SKI
- 1934 EDO 4650 FLOATS
- TRAINING AND DELIVERY



PRICE:

AIRFRAME - \$475K
ENGINE - \$75K
FLOATS - \$50K
SKIS - PRICELESS
= \$600,000

ENGINE: TTSOH: 145.80

AIRFRAME: TTSN: 6917



SERIOUS INQUIRIES,
PLEASE CONTACT
HSW@SHGLLC.COM.

TRAVELAIR-6000B-NC9084.COM



of our 10 human analysts. Historically, our analysts only look at flights that the aircraft owner or pilot specifically asks us to look at—perhaps because the engine ran rough or hiccupped or they’re asking for help diagnosing some other known issue.

In recent years, we’ve been using a specialized AI model called “FEVA2” to screen every uploaded flight for evidence that an exhaust valve might be burning. Project GADfly is far more ambitious. Its goal is to screen every uploaded flight for virtually anything that looks out of the ordinary, flagging any such flights for scrutiny by our human analysts. This should allow us to be a lot more proactive, enabling us to alert the aircraft owner to all sorts of mechanical or operational issues that he doesn’t yet know he has.

What seemed like a manageable task at its outset two and a half years ago proved far more challenging than we expected. But I’m happy to report that we’ve made tremendous progress. So, here’s a look

at where we’ve been, where we are, and where this project is headed.

A daunting challenge

When we first launched Project GADfly, we knew we had an impressive dataset to work with. Savvy’s database contains more than 9 million flights, each loaded with detailed engine monitor information—exhaust gas temperatures (EGTs), cylinder head temperatures (CHTs), fuel flow, rpm, manifold pressure, oil temperature, and a lot more. Modern engine monitors capture this data once per second. That’s 3,600 observations for each hour of flight time, each observation containing values from dozens of sensors.

This is truly big data, a rich resource for training a machine-learning AI. The challenge is that those 9 million flights in our database are “unlabeled”—that is, we don’t know which of the flights are normal and which contain anomalies. For machine learning, that’s a significant hurdle. It’s

hard to teach an AI to spot trouble without training it on lots of examples of what “normal” and “not normal” look like.

Asking our team of 10 expert data analysts to examine these 9 million flights—or even a small fraction of them—would not be feasible. There had to be a better way, but what? We quickly realized we needed expertise beyond our in-house capabilities to solve this.

Calling in the experts

Fortunately, we found the right collaborator in John Sipple, a machine learning and anomaly detection specialist at Google with a doctorate in computer science who also teaches in the computer science program at The George Washington University. His accomplishments included developing machine-learning anomaly detection systems for cybersecurity, counterfeit detection, agriculture, missile defense, and optimizing climate control systems in commercial buildings.

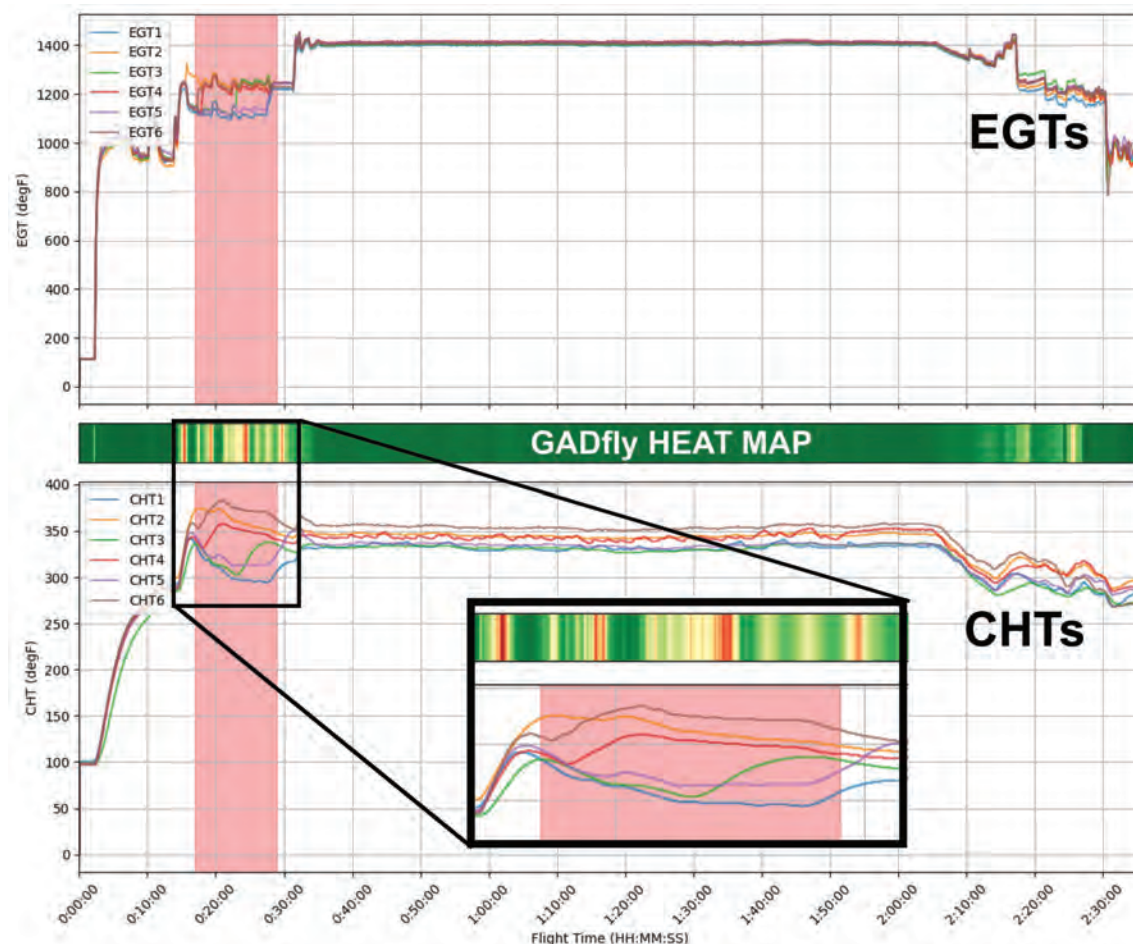


Figure 1
Odd-even CHT
split during climb.

As luck would have it, Sipple is also a pilot and owner of a Diamond DA40NG. So, when we explained what we were trying to do with Project GADfly, he agreed to help us, working with Savvy's analytics expert Adam Goler, a CFII with a doctorate in physics. Sipple also enlisted a talented researcher, Catherine Nguyen, to assist.

Their challenge was to develop a multilevel neural network "classifier"—a trainable AI that could examine engine monitor data from a flight and determine whether it appears normal or anomalous. Without labeled data, this was no small feat, but Sipple brought a solution to the table that he had developed: a technique called MADI, or Multivariate Anomaly Detection with Interpretability.

How MADI works

MADI is an ingenious approach to dealing with large volumes of unlabeled observations when the observations include a lot of variables and most of the observations are normal with abnormal ones being rare. MADI uses a trick called "negative sampling," and here's how it works.

First, we select several thousand actual flights from our database and arbitrarily label them as "normal" even though we know some small number of them are probably anomalous—perhaps due to some sort of mechanical problem or pilot mismanagement—but such anomalies are rare, so the labeling inaccuracy is minimal.

Next, we create several thousand phony flights populated with randomized values for EGTs, CHTs, rpm, manifold pressure, fuel flow, and other values, each randomized value limited to a range that is plausible. Almost all these phony flights would be highly anomalous, although there's a miniscule chance that one or two might be normal—about the same chance as 1,000 monkeys at typewriters producing a work of Shakespeare. We label all these phony flights as "anomalous," knowing that the labeling inaccuracy is negligible.

Now we can use this combination of real "normal" flights and phony "anomalous" ones as a labeled training set to train our neural network classifier on how to distinguish between normal and abnormal. The collection of normal flights

defines a multidimensional manifold (or "hyperblob" if you prefer) whose boundary defines what combinations of observed values are normal. The classifier determines whether a new observation falls inside or outside this boundary and generates an "anomaly score" that indicates how likely it is that this observation is anomalous. Voilà!

Real-world example

To get a better feeling for this, look at a few graphics of actual cases where the GADfly model has scanned flights and identified anomalies. Let's start with Figure 1. For this flight, the classifier identified an 11-minute anomalous segment during the climb phase. In this graphic, the top chart shows EGTs and the bottom chart shows CHTs. Between the two charts is a "heat map" showing the GADfly classifier's anomaly score for each one-second observation during the flight, depicting green for observations that seem normal, red for ones that seem highly anomalous, and yellow for ones that seem moderately anomalous.

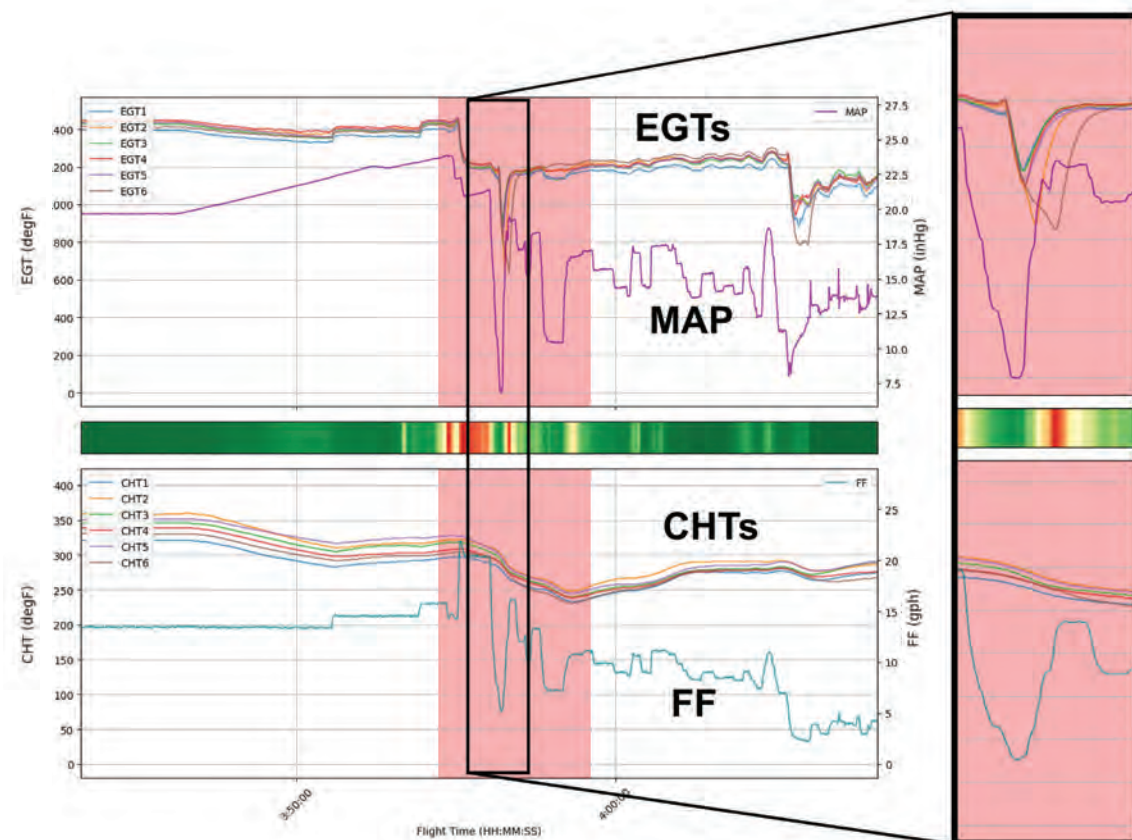


Figure 2
Brief fuel
starvation while
switching tanks.

If you look closely at the 11-minute segment highlighted in pink, you'll see that the EGTs look normal and stable, but there's a noticeable split in CHTs between odd and even cylinders. It's this CHT split that caught the attention of the GADfly classifier. But what caused it?

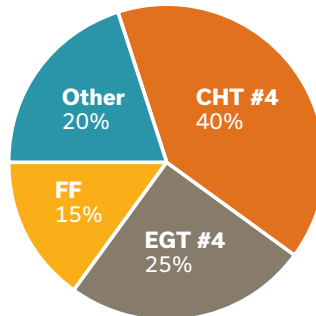
We had one of our human analysts take a look. There were two obvious things that could cause an odd/even CHT split: an induction leak on one side of the engine, or a cooling baffle problem on one side of the engine. An induction leak would also cause an EGT split, but we don't see that in the data. So, the analyst's presumptive diagnosis was that it was most likely a cooling baffle issue.

Sure enough, the owner's A&P found an improperly secured cooling baffle that was leaking under pressure. Securing the baffle caused the issue to go away.

More examples

Figure 2 shows another flight where the GADfly classifier identified a relatively brief anomaly nearly four hours into a

GADfly blame analysis



long cross-country flight. In the highlighted segment, note that the EGTs and fuel flow dropped suddenly before recovering, but was quite unstable after that. The heatmap turned red during the engine stumble. We asked the pilot about this, and he said the stumble occurred just after he switched tanks. He thought that perhaps he'd failed to get the fuel selector in the detent, but the data strongly suggested there was vapor in the fuel line from the newly selected tank.

Finally, Figure 3 shows a rather dramatic anomaly caught by the GADfly classifier. The pilot took off with a full-rich mixture, then leaned in the climb to maintain constant EGTs—a good technique—but ultimately might have overdone it, as cylinder number four was extremely unhappy. Note how EGT number four spiked downward and CHT number four headed rapidly toward redline. These are clear signs of heavy detonation or more likely preignition. The pilot obviously sensed something was seriously wrong because he quickly throttled back and put the airplane back on the ground. We suggested sticking a borescope into cylinder number four and inspecting for possible damage, especially to the piston.

Adding explainability

Automated detection of anomalies is great, but we really wanted GADfly to do more. Whenever it flagged a segment of a flight as anomalous, we also wanted it to tell why it did what it did—a concept known in AI as “explainability.” An automated alert is certainly useful, but understanding what

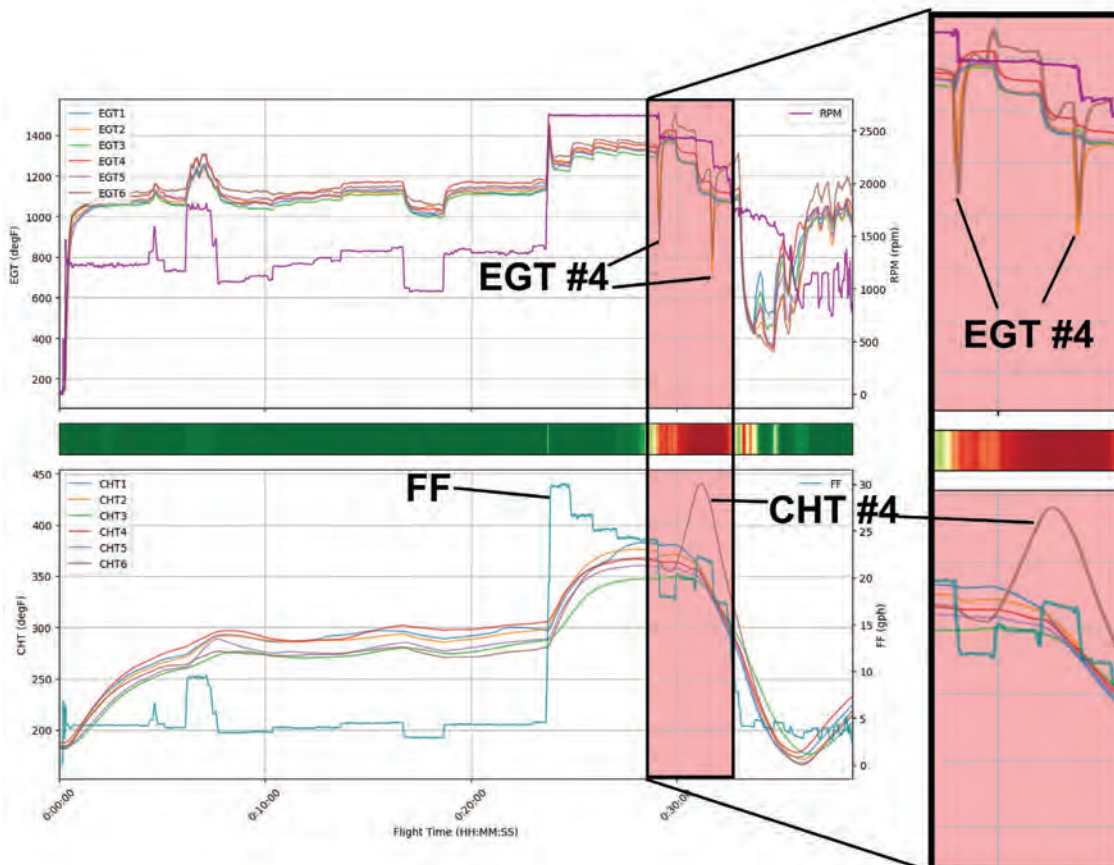


Figure 3
Cylinder number four preignition event during climb.

caused it would take GADfly to the next level and could be extremely helpful to our human analysts—both to develop a diagnosis and to prioritize which flights to look at first—and ultimately to mechanics tasked with troubleshooting an issue.

We've made some progress here. GADfly now provides a "blame analysis" (on left) that identifies which specific data elements—like a sudden EGT drop, CHT spike, or inappropriate fuel flow—contribute most to the anomaly score. Additionally, the model can group related anomalous observations with similar blame percentages into a single "event." Long-duration events, such as that 11-minute baffle problem, carry more weight than shorter ones, and very brief anomalies are often dismissed as false positives that can occur when the pilot does something dramatic like apply full power at takeoff or pull the mixture control back from rich of peak (ROP) to lean of peak (LOP) in cruise.

Right now, we're playing around with using GADfly's blame analysis as a prompt to a large language model (LLM) like ChatGPT or Grok to come up with a plain language explanation of why GADfly flagged a flight. We're even experimenting with training such an LLM in my book *Mike Busch on Engines* to see if we can get GADfly to produce a reasonable presumptive diagnosis. Are we having fun yet?

Looking ahead

Currently, GADfly is analyzing every Cirrus SR22 flight uploaded to Savvy's platform. We chose the SR22 because it's the most common aircraft in our database and has lots of sensors. The results are encouraging enough that we're about to begin training, tuning, and testing the model on other popular airplane models, starting with the Beech Bonanza. Our plan is ultimately to extend GADfly to many other makes and models for which

we have sufficient training data in our database. (But if you fly a Yak or a Vari-Viggen, don't hold your breath.)

Soon, we hope to start having engine data automatically transferred to our platform via telemetry so we can warn owners of GADfly-detected anomalies in near real time. Ultimately, we think it might be possible to put a GADfly classifier inside the engine monitor itself, providing truly real-time anomaly detection alerts in the cockpit. That sure would be cool! ■

mike.busch@savvyaviation.com

► savvyaviation.com

Ask the A&Ps /

Mike Busch, Paul New, and Colleen Sterling answer your toughest aviation maintenance questions on our *Ask the A&Ps* podcast. Submit questions to podcasts@aopa.org
► aopa.org/ask-the-a-and-ps

91



GOT RANGE ANXIETY? CHANGE THIS

Our sensors provide the gold standard of fuel level reliability with unparalleled accuracy.

Fly the skies with confidence today.

- ✓ FOR APPROVED AND EXPERIMENTAL AIRCRAFT
- ✓ COMPATIBLE WITH MOST MODERN GAUGES
- ✓ FAA AND EASA* APPROVED

SEE YOU AT
EAA AIRVENTURE OSHKOSH 2025
Booth # **3119**

CIESCORP INC.

SCAN TO LEARN MORE
CIESCORP.COM

*EASA approval through our partnership with Aerospace Logic