augmenting biologging with supervised machine learning to understand *in situ* behavior of *Chrysaora fuscescens*
biologging: a window into *in situ* behavior

Photo by Jurgen Freund

Photo by Doug Nowacek

Mooney & Katija et al. (2015)
the tether method for tag retrieval

Hays et al. (2008)
Fossette et al. (2015)
Fossette et al. (2016)
motivating questions

1. Can we extract meaningful *in situ* data from suboptimal biologging methods?

2. How can we use motion sensor data to study different *in situ* behaviors?
motivating questions

1. Can we extract meaningful *in situ* data from suboptimal biologging methods?

2. How can we use motion sensor data to study different *in situ* behaviors?

Both questions can be formulated as supervised machine learning problems.

\[ y = f_\theta(x) \]

- \( x \): inputs (features of motion sensor data)
- \( y \): outputs (labels of behavior or tether influence)
- \( f \): function (from some class of functions)
- \( \theta \): parameters of function
**in situ experimental setup**

SPOT drifter

3-7 m drogue

30 m monofilament

ITAG

optimistic scientists

100 Hz accelerometer

gyroscope

magnetometer

pressure

BlueROV2 + GoPro HERO5 Black
footage annotation

Each second of footage annotated with **tether status:**

<table>
<thead>
<tr>
<th>influenced</th>
<th>uninfluenced</th>
<th>unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="taut tether" /></td>
<td><img src="image2" alt="slack tether" /></td>
<td><img src="image3" alt="unknown" /></td>
</tr>
</tbody>
</table>

and **behavior:**

<table>
<thead>
<tr>
<th>swimming</th>
<th>drifting</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4" alt="swimming" /></td>
<td><img src="image5" alt="drifting" /></td>
</tr>
</tbody>
</table>
uninfluenced swimming
tether-influenced swimming
data preprocessing

high-pass filtering (FIR, 0.2 Hz cutoff)

motion sensor + annotated footage alignment

swimming

uninfluenced

tether unknown

segmentation + featurization

spectrum of dynamic acceleration

representative pulse length

2.05 s

feature 1

feature 2

…

feature 56

chosen using sequential forward selection
quadratic discriminant analysis (QDA)

**generative model:**
\[ y \sim \text{Bernoulli}(\phi) \]
\[ x \mid y = 0 \sim \mathcal{N}(\mu_0, \Sigma_0) \]
\[ x \mid y = 1 \sim \mathcal{N}(\mu_1, \Sigma_1) \]

\[ \phi \in [0, 1] \]
\[ \mu_0, \mu_1 \in \mathbb{R}^p \]
\[ \Sigma_0, \Sigma_1 \in \mathbb{S}^+_p \]

learned via maximum likelihood estimation

**prediction:**
\[ f_{\phi,\mu_0,\mu_1,\Sigma_0,\Sigma_1}(x) = \arg \max_{y \in \{0,1\}} p(y \mid x) \]
in situ deployments

Average per animal:
4.5 hours of tag data
14.5 minutes of video footage

May 14, 2018

May 17, 2018

April 24, 2018

deployment start
tag is recording
tag retrieval (next day)
classifier cross-validation accuracies

tether-influence classifier
97.6% (standard error = 0.24%)

activity (swim vs. drift) classifier
99.5% (standard error = 0.1%)
how should jellyfish orientation be described?
a fine-scale portrait of \textit{in situ} behavior over 5.5 hours
a fine-scale portrait of *in situ* behavior over 5.5 hours
a fine-scale portrait of in situ behavior over 5.5 hours
a fine-scale portrait of *in situ* behavior over 5.5 hours
a fine-scale portrait of *in situ* behavior over 5.5 hours
a fine-scale portrait of \textit{in situ} behavior over 5.5 hours
a fine-scale portrait of *in situ* behavior over 5.5 hours
a fine-scale portrait of *in situ* behavior over 5.5 hours
In situ behavior predicted as tether-influenced

In aggregate, behavior was predicted to be tether-influenced **11.4%** of the time.
how does tether influence impact the behavioral data?

Orientation change induced by tether influence is **significantly greater** than orientation change due to normal turning behavior.

Our tether-influence classifier captures this artificial effect on behavioral data, and can be used to remove affected portions.
In aggregate, animals were predicted to be swimming 99.3% of the time.

Costello, Klos, & Ford (1998)
conclusions & future work

Supervised machine learning methods can help 1) enhance the quality of, and 2) interpret challenging data collected via suboptimal biologging methods.

Using these methods, we 1) quantified and removed the effects of the tether method, and 2) estimated that *C. fuscescens* spends the vast majority (~99.3%) of the time actively swimming.

These methods can help us investigate complex ecological and behavioral questions. Our approach is broadly applicable to biologging, and particularly promising for the challenging study of marine invertebrates *in situ*. 
many thanks to...

ITAG Development
David Mann
Aran Mooney
Alex Shorter

MBARI
Jared Figurski
Joost Daniels
Steve Haddock
Rob Sherlock
Chad Kecy
Brett Hobson
Dale Graves
Larry Bird
Johnny Ferreira
Paul McGill
Chris Wahl

Hopkins Marine Station
Will Gough
James Fahlbusch
Dave Cade
Jeremy Goldbogen

the David & Lucile Packard Foundation

Award #1455501