

# Parameterization Model

This analysis extracts swimming parameters from each of the 127 different experimental conditions tested. Then a random forest model is built and saved so that someone downstream can pull the full set of parameters (or whatever they need) given an arbitrary set of environmental conditions.

The experimental conditons are: flow rate, chlorophyll concentration, guano presence, and light presence.

The swimming parameters are: - Mean velocity magnitude - Velocity magnitude variance - Velocity magnitude autocorrelation slope, intercept, residual - Velocity magnitude bimodal test significance - Mean angle (horizontal/vertical)... plus same relative to flow direction - Angle coefficient of dispersion (horizontal/vertical) ... plus same relative to flow direction - Heading (horizontal) autocorrelation slope, intercept, residual - ... (there will be more)

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2

## -- Attaching packages ----- tidymodels 1.1.0 --

## v broom          1.0.5      v rsample          1.1.1
## v dials          1.2.0      v tibble          3.2.1
## v dplyr          1.1.3      v tidyr           1.3.0
## v infer          1.0.4      v tune            1.1.1
## v modeldata      1.1.0      v workflows       1.1.3
## v parsnip        1.1.0      v workflowsets    1.0.1
## v purrr          1.0.2      v yardstick       1.2.0
## v recipes        1.0.6

## -- Conflicts ----- tidymodels_conflicts() --
## x purrr::discard() masks scales::discard()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
## x recipes::step()   masks stats::step()
## * Use tidymodels_prefer() to resolve common conflicts.

## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:dplyr':
##
##   combine

## The following object is masked from 'package:ggplot2':
##
##   margin

##
## Attaching package: 'matrixStats'
```

```
## The following object is masked from 'package:dplyr':  
##  
##      count
```

```
## Warning: NAs introduced by coercion
```

The experimental conditions tested were: - Flow rates: 0, 0.6, 3.0, 5.9, 8.9 - Chlorophyll concentrations: 0.00, 1.01, 2.70, 2.80, 3.00, 3.70, 4.30, 4.50, 4.60, 5.50, 5.70, 5.90, 6.00, 6.10, 7.60, 9.00, 13.50, 16.00, 16.90, 19.00 - Guano presence: absent, present - Light presence: absent, present Of the possible combinations, 127 combinations were tested in an experimental setting.

Now we need to extract all of the swimming parameters for each experiment.

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):  
## collapsing to unique 'x' values
```

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):  
## collapsing to unique 'x' values
```

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):  
## collapsing to unique 'x' values
```