Predicting the settlement cycle and intensity of mussel larvae with GAM for real and functional data

Isabel Fuentes-Santos
U. Labarta, X.A. Álvarez-Salgado, M.J. Fernández-Reiriz

Instituto de Investigaciones Marinas
PROINSA Mussel Farm lab

Santiago de Compostela, July 2, 2016
Outline

1 Introduction

2 Dataset

3 Predicting settlement occurrence

4 Prediction settlement intensity

5 Conclusions
1 Introduction

2 Dataset

3 Predicting settlement occurrence

4 Prediction settlement intensity

5 Conclusions
Mussel raft culture is the main aquaculture industry in Galicia.

- Production above 260,000 Tn/year.
- 2nd producer in the world.
- Seed collection is crucial to support mussel aquaculture.
Seed collection

Life cycle of mussels

- Embryo
- Larval Stage
- Juvenile
- Adult

PELAGIC

Settlement

BENTHIC
Seed collection

Life cycle of mussels

Rocky shores

Collector ropes
Objective

- Forecast the timing and intensity of mussel larval settlement in the Ria de Ares-Betanzos, according with meteorological conditions.
Objective

- Forecast the timing and intensity of mussel larval settlement in the Ria de Ares-Betanzos, according with meteorological conditions.
I. Introduction

II. Dataset

III. Predicting settlement occurrence

IV. Prediction settlement intensity

V. Conclusions
**Response**: larval settlement on collector ropes at 1m.

**Weekly samplings from 2009.**

- **N (ind/m)**: settlement intensity.
- **Y**: larval occurrence ($Y = 1$ if $N > 500$)

**Environmental factors**: daily records

- **R (MJ/(m$^2$day))**: solar irradiance at CIS-Ferrol (Meteogalicia).
- **Q (m$^3$/s)**: sum of discharges from rivers Eume (Endesa S.A.) and Mandeo (Augas de Galicia).
- Wind speed, **W (m/s)**, and direction, **Θ**, at Vilano Buoy (Puertos del Estado)
Settlement patterns

Large settlement period with mid-spring peak.

Short settlement period.
Settlement patterns

- Large settlement period with mid-spring peak.
- Short settlement period.
1 Introduction

2 Dataset

3 Predicting settlement occurrence

4 Prediction settlement intensity

5 Conclusions
Model for settlement timing

- GAM with binomial family and logit link.
- Explanatory variables
  - $R$ mean solar irradiance recorded 30-45 days prior to sampling.
  - $Q$ mean river discharges recorded 10-30 days prior to sampling.
  - $W$ and $\theta$ mean wind speed and direction 10-30 days prior to sampling.
Model for settlement timing

- GAM with binomial family and logit link.
- Explanatory variables
  - $R$ mean solar irradiance recorded 30-45 days prior to sampling.
  - $Q$ mean river discharges recorded 10-30 days prior to sampling.
  - $W$ and $\theta$ mean wind speed and direction 10-30 days prior to sampling.

\[
P(Y = 1|R) = H(\alpha + f(R)) = \frac{\exp(\alpha + f(R))}{1 + \exp(\alpha + f(R))}
\]

---


Predictions for 2014

\[ Y \sim R + W \] selected by the algorithm.

Low probability of misclassification with both models.

\[ Y \sim R. \]

Prediction lag: 1-month

\[ R \approx 11MJ/m^2/day, \]

threshold for settlement occurrence.

I. Fuentes-Santos (IIM-CSIC)
Predictions for 2014

- $Y \sim R + W \Theta$ selected by the algorithm
- Low probability of missclassification with both models.
- $Y \sim R$.

- Prediction lag: 1-month
- $R \approx 11 \text{MJ}/(m^2 \text{day})$, threshold for settlement occurrence.
Does not detect the settlement reduction after the first episode.

\( R \approx 11 \text{MJ/}(\text{m}^2 \text{day}) \), threshold for settlement occurrence.
1 Introduction

2 Dataset

3 Predicting settlement occurrence

4 Prediction settlement intensity

5 Conclusions
Model for settlement intensity

- **GAM:** Two-stage zero-inflated Poisson model (ziplss family).
  1. **Larval occurrence:** 30-45 days solar irradiance (R)
  2. **Settlement intensity:** solar irradiance (R), river discharges (Q) and wind regime (W, Θ).

  \[
  E[N | fR, fQ] = H(\beta + g_1(fR) + g_2(fQ))
  \]

Model for settlement intensity

**GAM**: two-stage zero-inflated Poisson model (ziplss family).

1. **Larval occurrence**: 30-45 days solar irradiance ($R$)
2. **Settlement intensity**: solar irradiance ($R$), river discharges ($Q$) and wind regime ($W, \Theta$).

**FGKAM** (Functional generalized kernel additive model).

- Negative binomial family.
- **Explanatory variables**: monthly curves (15-45 days before sampling)
  - $f_R$: solar irradiance
  - $f_Q$: river discharges
  - $f_W, f_\Theta$: wind speed and direction.
Model for settlement intensity

- **GAM**: two-stage zero-inflated Poisson model (ziplss family).
  1. **Larval occurrence**: 30-45 days solar irradiance ($R$)
  2. **Settlement intensity**: solar irradiance ($R$), river discharges ($Q$) and wind regime ($W$, $\Theta$).

- **FGKAM** (Functional generalized kernel additive model).
  - Negative binomial family.
  - **Explanatory variables**: monthly curves (15-45 days before sampling)
    - $f_R$: solar irradiance
    - $f_Q$: river discharges
    - $f_W$, $f_{\Theta}$ wind speed and direction.

Model for settlement intensity

- **GAM**: two-stage zero-inflated Poisson model (ziplss family).
  
  1. **Larval occurrence**: 30-45 days solar irradiance ($R$)
  2. **Settlement intensity**: solar irradiance ($R$), river discharges ($Q$) and wind regime ($W, \Theta$).

- **FGKAM** (Functional generalized kernel additive model).
  
  - Negative binomial family.
  - **Explanatory variables**: monthly curves (15-45 days before sampling)
    
    - $f_R$: solar irradiance
    - $f_Q$: river discharges
    - $f_W, f_\Theta$: wind speed and direction.


$$E [N|f_R, f_Q] = H (\beta + g_1(f_R) + g_2(f_Q))$$

---

Prediction 2009-2014

Better fit using functional data (FGKAM).

I. Fuentes-Santos (IIM-CSIC)

GAM and FKGAM to predict larval settlement
Better fit using functional data (FGKAM).
Prediction for 2015

Good prediction of the settlement cycle
Tend to underestimate settlement peaks

I. Fuentes-Santos (IIM-CSIC)

GAM and FKGAM to predict larval settlement
Good prediction of the settlement cycle
Tend to underestimate settlement peaks
Conclusions

- **Solar irradiance** influences larval production and development.

- **GAM**: cost-effective model to predict the onset and end of the settlement period **1 month in advance**

- **Functional GAM**
  - Take advantage of the information provided by continuous monitoring of meteorological and environmental conditions.
  - Cost-effective model to predict settlement timing and intensity **2 weeks in advance**.

- Helpful to schedule seed-collection.

---

THANK YOU
Collect collector ropes covered with Scotch-Brite scouring pads are kept in seawater during 30 days to favor the development of an adequate biofilm. Seawater is filtered every two days to prevent the attachment of epifauna. Collector ropes are deployed in a long-line in Arnela. After a week ropes are substituted by new ones.

**Sampling**
- Collect 3 sections of known area ($12cm^2$) of the scouring pad attached to each rope.
- Preserve samples on 70% ethanol until processing

**Laboratory processing**
- Detach settled individuals from the pads.
- Count individuals in a binocular microscope.