Informing Spatio-temporal Correlation in Surf Smelt Egg Detection to Improve HPA Protection of Forage Fish Spawning Beaches

Timothy Quinn, Kirk Krueger, Ilai Keren

Background

The Washington Department of Fish and Wildlife (Department) administers the Hydraulic Project Approval (HPA) program under authority granted in the state Hydraulic Code. The intent of the HPA program is to protect all fish life by regulating the construction of hydraulic projects or performance of other work that uses, diverts, obstructs, or changes the natural flow or bed of any of the salt or fresh waters of the state. These regulations are intended to reflect the best available science and practices related to protection of fish life. In addition, the Department is expected to incorporate new information as it becomes available so long as that information allows for alternatives that provide equal or greater protection for fish life. The intent of this document is to describe how the Department is incorporating new information to better protect forage fish spawning beaches in Puget Sound.

The Department conditions HPA permits to protect marine beach spawning fish, principally Surf Smelt and Sand Lance, during construction of shoreline armoring, overwater structures (dock, piers and floats) and other related activities. These conditions, which typically apply to beaches previously demonstrated by the Department or its certified partners to support forage fish spawning, commonly establish construction work timing windows to protect eggs during the spawning season. Information documenting the presence of forage fish eggs on the 4023 km (2500 miles) of marine beaches throughout Puget Sound is based on a relatively long (~30 year) but sparse sampling effort. This incomplete sampling effort reflects the scientific uncertainty related to forage fish ecology, the intensive nature and logistical difficulties of sampling (see Quinn et al. 2012) and the sheer length of Puget Sound shorelines.

A single forage fish egg survey consists of sampling upper intertidal beach sediments along a transect (parallel with the water line) 30 m (100 ft) in length. At four locations along this transect, approximately 500 ml of surface sediment is collected. The sediment subsamples are combined into a single sample and taken back to the lab for processing (http://wdfw.wa.gov/publications/01209/wdfw01209.pdf). When a survey sample contains forage fish eggs, the Department applies an occupancy standard (D. Pentilla, personal communication) that establishes the following conditions: The occupied status of a single 30 m transect is extended along the beach in both directions up to a distance of 152 m (500 ft) or until there is a change in beach type based on field observations. The standard based on best professional judgment (D. Pentilla personal communication) establishes the maximum spatial scale of an occupied beach (2 * 152 m = 305 m) based on a single occupied transect. Construction activities anywhere within this occupied beach are conditioned in HPAs as if the beach were occupied during the forage fish spawning seasons.

While this standard addresses occupied beaches based on the presence of eggs found during sampling, it does not address shoreline in “unknown occupancy status”. Much of the remaining shoreline in unknown occupancy status appears to contain suitable habitat based on sediment size distributions. In other words, based on our current knowledge of habitat use by Surf Smelt, many beaches that appear suitable for forage fish spawning are either under-sampled such that the Department has not been confident in surveys to find eggs when eggs are present, or unsampled (see Quinn et al. 2012). This issue becomes particularly apparent when an occupied beach boundary based on the occupancy standard ends along a continuous uninterrupted beach type that appears suitable for spawning and for which sampling has not been conducted.

Among the many outstanding questions related to forage fish spawning behavior and habitat use, one question is repeatedly raised by Department biologists during the permitting process. That questions is: What is the spatial extent of an occupied beach?, or put another way, If forage fish eggs were detected at a single 30 m transect, then how far, on average, from that transect are eggs likely to be found? This question has taken on new urgency because Department permits are increasingly appealed by third party interests, who point out that the Department is issuing permits with no provisions for forage fish protection on beaches that appear capable of supporting forage fish spawning, are un- or under-sampled, and are often in close proximity to forage fish occupied beaches.

Incorporating New Information

We, (Habitat Science Division) initiated studies in 2014 that will address the issue of beach occupancy described above. However, this work is extensive and will require years of study and likely additional resources. In the meantime, we used a study by Quinn et al (2012) in an attempt to better inform the process we currently use to designate the scale of an occupied beach based on an occupied transect described above. In that study we surveyed 51 Camano Island beach sites for forage fish eggs twice per month over the course of a year. At each site visit, we a conducted a forage fish egg survey described above. The intent of that study was to: 1) Document the annual timing of Surf Smelt and Sand Lance spawning, 2) Explore the relationship between beach and backshore habitat characteristics and egg counts, and 3) Better understand the variability in abundance and detection of eggs among sites. Beach sites on Camano Is. were established in a randomly stratified manner wherein, we randomly establish an initial site on the beach and placed other sites at fairly regularly spaced intervals approximately 1.25 km in length over all beaches that represented potential spawning habitat.

For each Camano site visit we estimated the number of live and dead eggs, and used these data to explore patterns (build explanatory models) of presence/absence information. In particular, we were interested to know if site scale characteristics found to be important in models explaining egg abundance (Quinn et al. 2012) were important in explaining presence. In addition, we were interested in knowing if egg presence/absence among sites was related to season and if egg presence at sites were spatially autocorrelated, that is, if egg presence at one beach site during one sample session was related to or could be predicted from the presence of eggs at another site on that same sample session.

Model input data included presence/absence results of 24 surveys taken at two week intervals over the course of a year for each of 51 sampling sites, and four covariates, i.e., physical characteristic, measured once at each site. Covariates included Northness, which is a measure of aspect converted to a standardized polar coordinate system that produces values ranging from -1 (South) to 1 (North), Global Site Factor, which is defined as the proportion of direct and indirect global solar radiation at a given site relative to an open and unshaded location. Mean Maximum Beach Temperature, and Fetch. As outlined in Quinn et al. (2012), all four of these covariates had some support for inclusion in models that helped explain annual egg abundance and thus were considered potentially important here. However, of the four potential covariates only Northness and GFS showed any potential effect on presence/absence and thus were included in further analyses.

We took a Bayesian approach to this analysis for several reasons, some more technical than others such as flexibility in hierarchical models, and the fact that imputation of the missing values is “automated". Others more theoretical reasons include the ability to interpret results as a probability of finding eggs on any given beach to aid in future management. Diffuse (non-informative) priors were put on all hyper-parameters.

A cosine function was used to model the seasonality of spawning:

$$P(y_{it} \geq 1) = logit^{-1}(\mu_i + R\cos(\vartheta + 2\pi ft)) \quad (1)$$

Where y_{it} are observed outcomes of a *Bernoulli random variable* denoting success of finding eggs at site *i* and time *t*. μ_i is the mean rate of success for the site on the logit scale. *R* and ϑ are the amplitude and phase of the seasonal trend, respectively, and *f* is the frequency (24/26.5) of visits/total bi-monthly sampling periods in a year.

A first order autoregressive term was used to model the correlation between mean rate (presence) of sites one lag (~1.25 km) apart.

$$\mu_i = (1 - \Phi)X\beta + \Phi\, \mu_{i-1+}\varepsilon_i \quad (2)$$

Where $X\beta$ are common effects of covariates and Φ is the correlation between mean rate of sites one lag apart and $\varepsilon \sim N(0, \sigma^2)$ is individual level site to site variation.

Model parameters were estimated via Gibbs sampling (MCMC) in JAGS 3.4.0. (Plummer, 2003) All models consist of 1200 independent draws from 3 parallel chains initiated at overdispersed values. After discarding the first 1,000 iterations as burn in, chains were run for 16,000 iterations and thinned every 40. Convergence and sample independence were assessed visually and by ensuring an expected value of $\widehat{R} < 1.1$ and effective sample size > 1,100 for all parameters (Gelman and Rubin, 1992).

Four models fit to the data included 51 rates of success (for each site) explained by northness and GSF: the Null model included no other parameters, a seasonal trend only model (Season), an autoregressive only model (AR(1)), and a full model that includes both the seasonal and autoregressive terms (Full). DIC (deviance information criteria) based model selection clearly favored models with both a seasonal trend and auto-correlation:

Table 1. Reduction in deviance associated with the four candidate models. Each model included northness and GFS as covariates.

Model	DIC
Null	1036
Season	704
AR(1)	1013
Full	686

In a posterior predictive check the full and seasonal models correctly classified a site at any point in time as eggs present or absent with 82% success rate (Fig. 2) while the null and AR(1) model had successfully classified sites 71% of the time.

Results

Correlation between sites was high with median expected values of Φ = 0.924 and 95% highest posterior density interval = (0.816, 0.999, Fig. 1). The inclusion of a correlation coefficient contributed to smoothing some of the differences between sites.

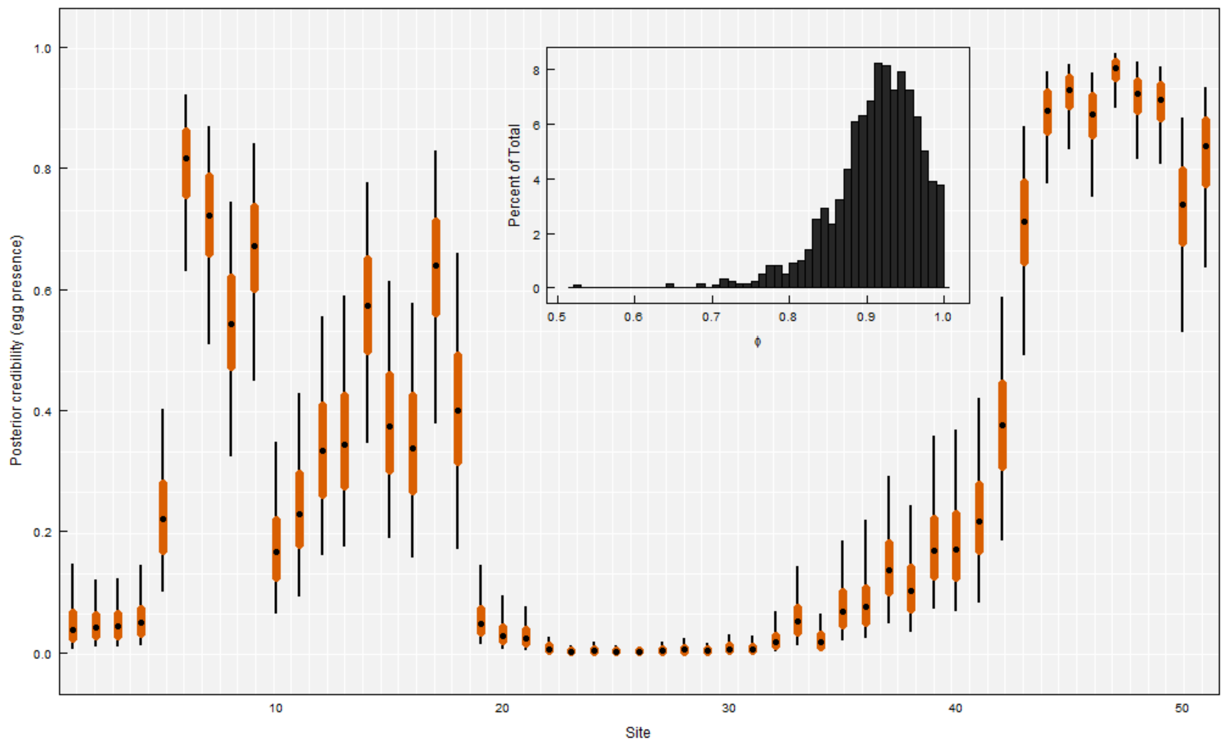

Figure 1. Bayesian 95% credible intervals (black), quartiles (orange) and medians for posterior presence at each site from the full model. Values are based on 1200 independent MCMC samples. Insert is the histogram for the correlation coefficient between adjacent sites (AR1) from the same model.

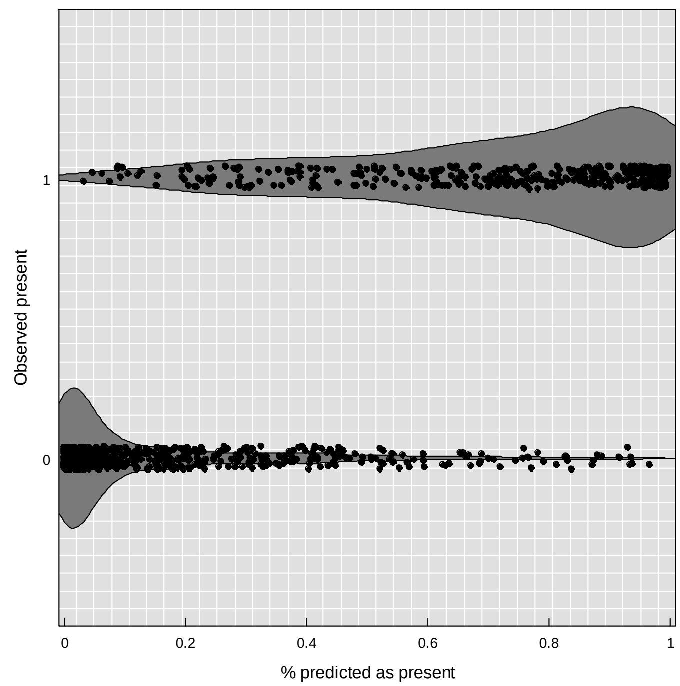

Figure 2 Posterior predictive check. A vector of 1200 random draws was generated from the posterior $[y^* \sim Ber\left(\hat{\theta} = \Pr(\theta | y)\right) \in \{0,1\}]$ for every point (site x time) in the data set and the proportion of successes (from 1200) plotted against the true value of Y_{it}.

Implications for Management

The current standard in 2016 for delineating an occupied length of beach used by spawning Pacific Sand Lance and Surf Smelt based on a single transect is 305 m. Based on the work described above, the Department could consider extending the previous standard of beach occupancy, based on an occupied forage fish sampling point (i.e., the center of a 30 m sampling transect) from the current 305 m to a distance of 1267 m, where 1267 m is the median distance (interquartile range = 70 m) between the 51 sites on Camano Islands. In other words, beach occupancy for a single occupied forage fish sampling point (center of the 30 m sampling transect) could be extended along the beach in both directions up to a distance of 634 m or until there is a change in beach type (historically based on Washington Department of Natural Resources ShoreZone inventory data or field survey). Under this scenario, the new standard would replace the old standard for establishing the spatial scale of an occupied beach based on a single occupied transect. Importantly, the result of this analysis is limited to Surf Smelt. Pacific Sand Lance spawning locations in the Camano Island were too sparse to analyze for spatial autocorrelation.

Literature Cited

Gelman, A. and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7: 457-511.

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003): 1-10.

Quinn, T. Krueger, K. Pierce, K. Penttila, D. Perry, K. Hicks, T. and D. Lowry. 2012. Patterns of Surf Smelt, *Hypomesus pretiosus*, intertidal spawning habitat use in Puget Sound, Washington State. Estuaries and Coasts 35: 1214-1228.