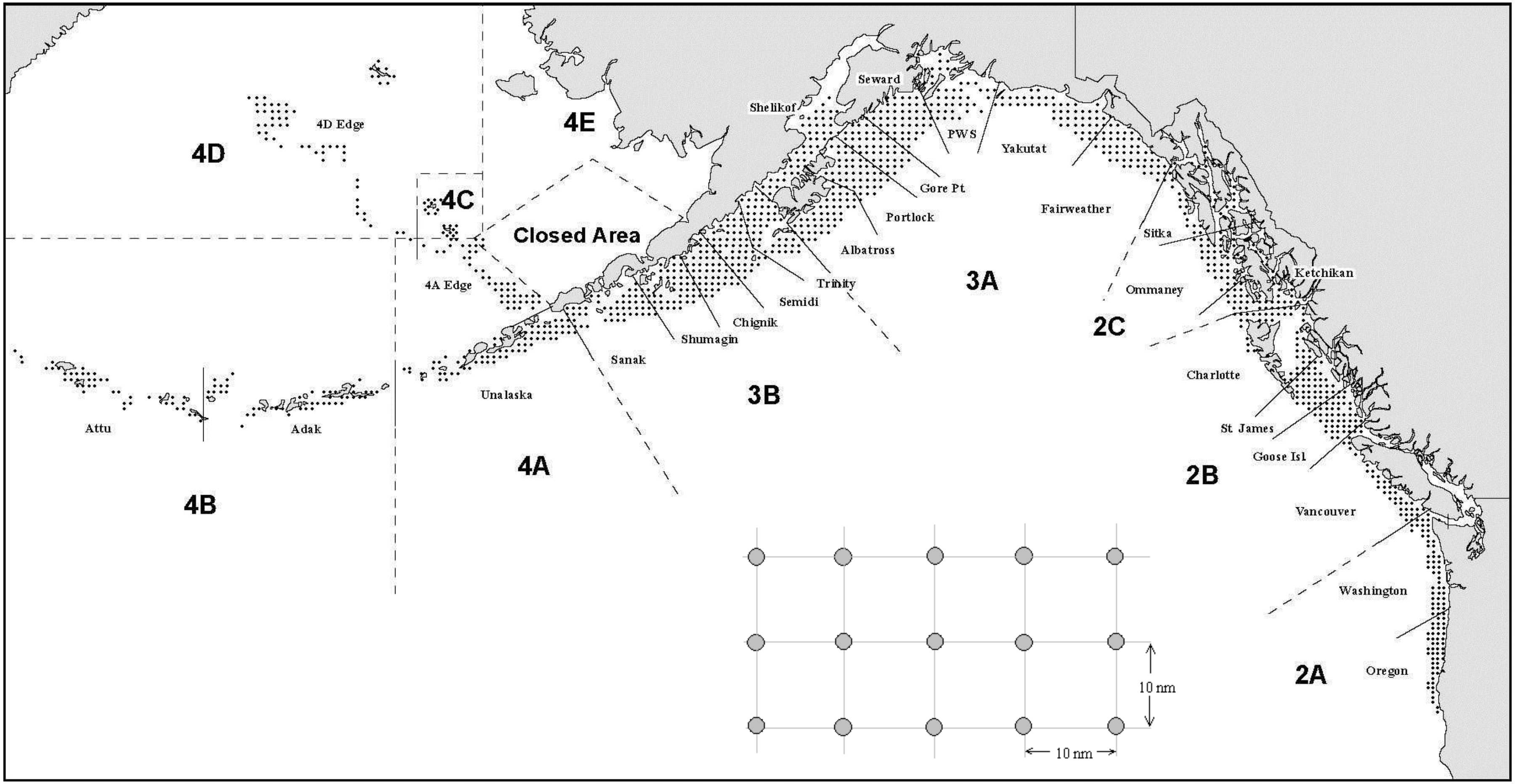
Spatio-temporal modelling of setline survey data

PART 1: Background and methods

Background

- IPHC standard setline survey
  - Annual survey
  - Over 1200 standard stations
  - On grid with 10 nmi spacing since 1998



Background

- Not all regions surveyed annually
  - Some regulatory areas not surveyed annually in early years of survey, e.g.,
    - eastern Bering Sea (Area 4CDE) only surveyed twice (2006, 2015)
    - northern California (2013, 2014) and Salish Sea (2011, 2014) only surveyed twice
  - Deep, shallow waters and other coverage gaps part of current piece-wise annual expansion program (2014-2019)

Survey WPUE and NPUE

- Mean survey O32 WPUE and NPUE are used as an indices of halibut density:
  - Index density in each regulatory area
  - Weighted by bottom area to create coastwide indices
- WPUE is used for apportionment, while NPUE is used in the stock assessment.

Survey WPUE and NPUE

- At present, various methods and data sources are used to account for incomplete spatial coverage:
  - NMFS longline sablefish survey indexes deep water (>275 ftm) in Areas 2B, 2C, 3A and 3B
  - NMFS West Coast trawl survey used to index extreme southern part of range in California
  - In areas that have had a survey expansion, ratio of WPUE at all stations to that at standard grid stations used as a scaling factor in years with no expansion
  - Otherwise, setline survey 20-275 ftm data assumed to be representative of 0-400 ftm
- Further adjustments are made to account for competition for baits and for the timing of the survey relative to the harvest.

Spatio-temporal modelling

- Spatial dependence:
  - Halibut distribution is not random (and neither is survey station placement!)
  - There are areas of high and low density, so, for example, a survey station with high WPUE (or NPUE) is more likely to be near other stations with high WPUE than stations with low WPUE
- Temporal dependence:
  - Patches of high and low density persist over time

Goals

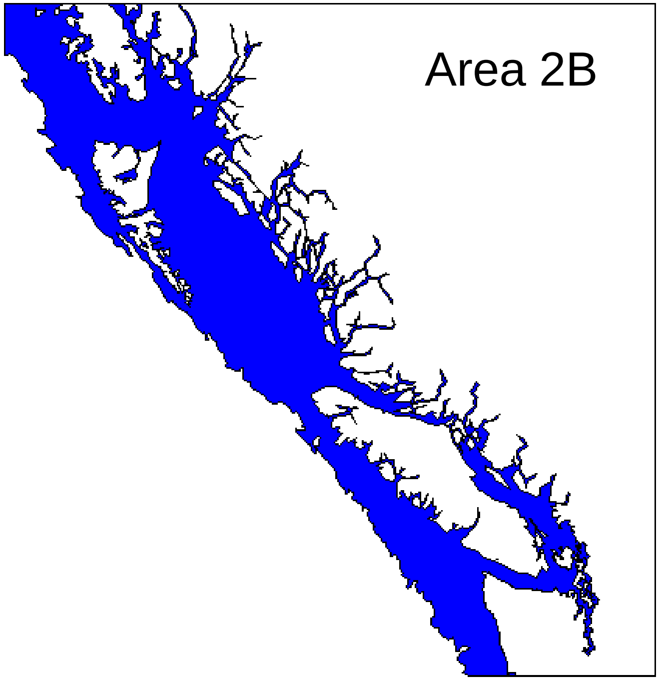
- Monitor changes in spatial distribution of halibut over time.
- Model relationship between density and covariates (while accounting for spatial dependence).
- **Improve WPUE and NPUE indices of density:**
  - Predict WPUE and NPUE at unsurveyed stations (coverage gaps, ineffective stations) and avoid need for numerous spatial coverage adjustments
  - Reduce effect of random variation in WPUE index
  - Account for uncertainty due to coverage gaps
- Help plan frequency of future expansions and optimise allocation of survey effort.

INLA

- Integrated nested Laplace approximation
- Computationally efficient
  - Our past attempts at spatio-temporal modelling of survey data (2008-2010) using MCMC were extremely slow
- Works on non-convex domains, such as ocean regions with rugged coastlines and islands
  - Not necessarily true of standard geostatistical methods
- Has a fairly flexible, easy to use R package, with good support from the developers.
- Approach used successfully for fisheries data by other scientists (e.g., Jim Thorson)

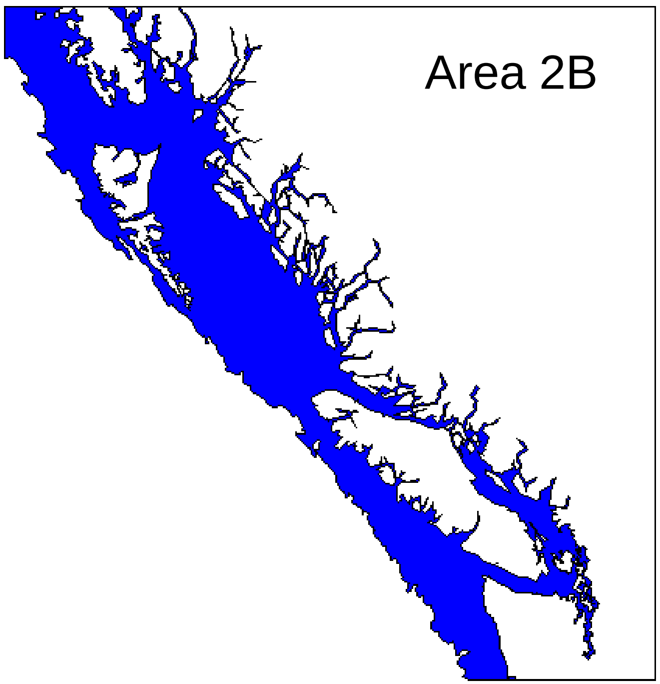
The domain of interest

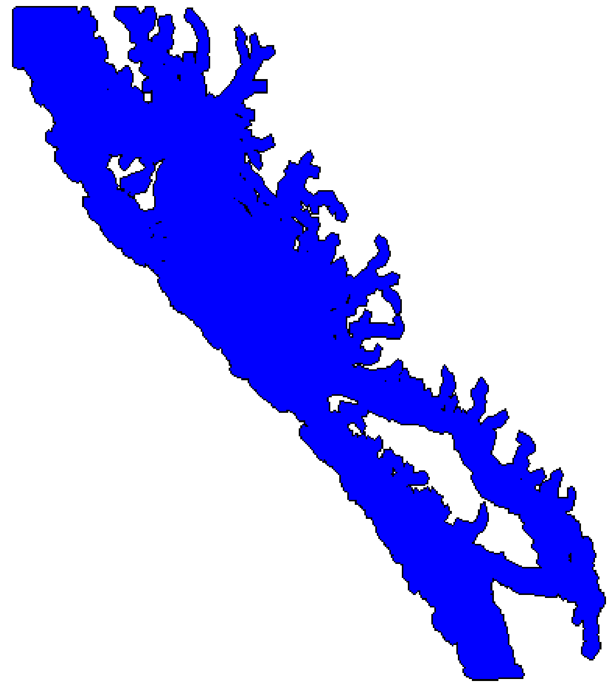
- Begin by defining a boundary for the domain of interest. 0-400 fathoms seems obvious…



The domain of interest

- …but to avoid boundary effects (high variance near boundary), better to extend domain slightly:





Mesh construction

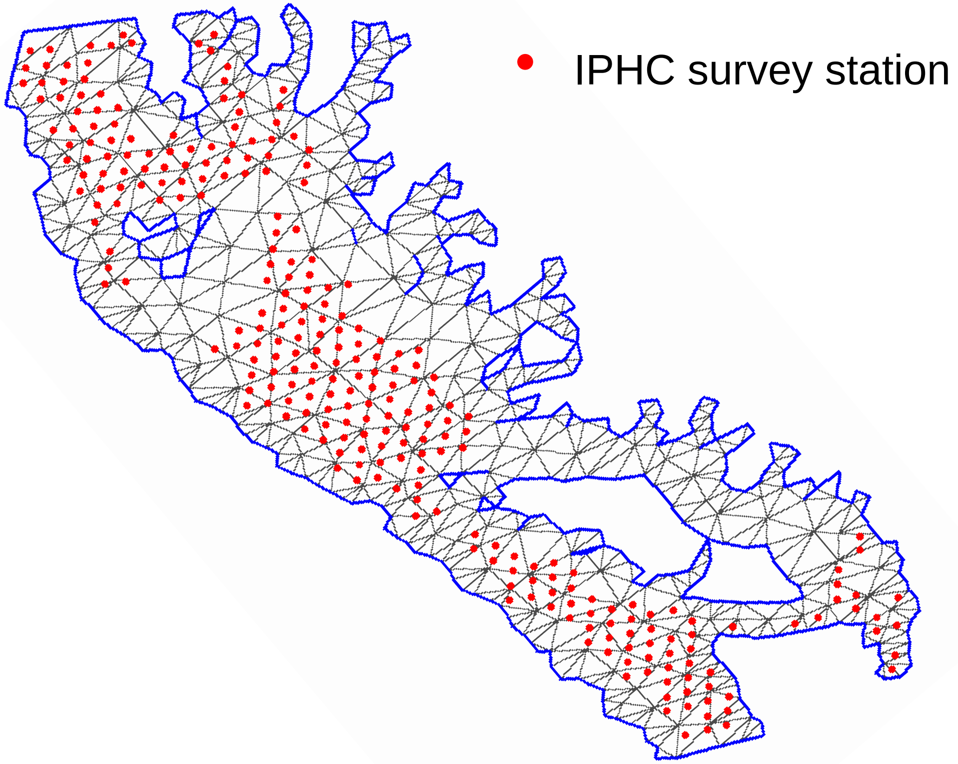
- INLA uses a set of basis functions defined on a triangulated mesh – need to define the mesh!

Mesh definition:

- Provides comprehensive spatial coverage
- Not so complex as to make model fitting prohibitively slow

Selecting the number of vertices involved some trial and error

- ~500 vertices used for each regulatory area to date



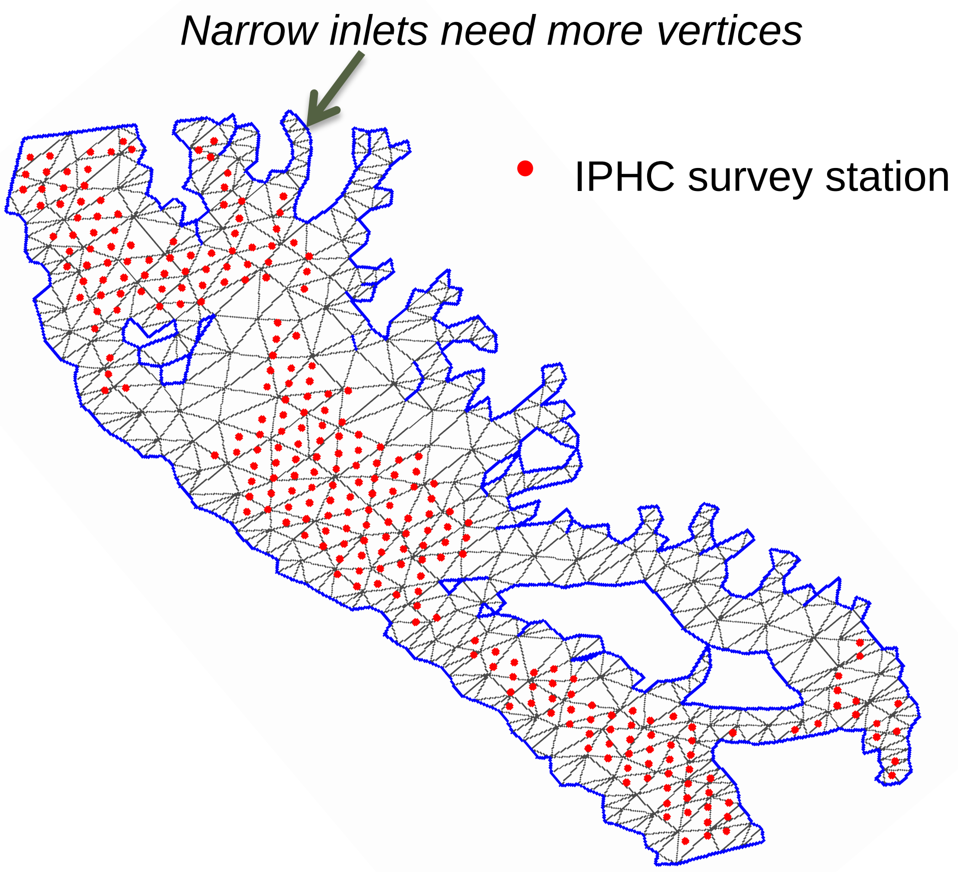
Mesh construction

- The advantages of expanding the boundary:

Expanding narrow inside-waters reduced the required complexity of the mesh.

The boundary was further smoothed as part of the mesh construction algorithm.

Note also that the domain of interest extends beyond the regulatory area boundaries.



Statistical models

- Station-level WPUE data:
  - Skewed distribution
  - In some areas, high probability of stations with WPUE or NPUE of zero (no halibut caught)
- For WPUE data, we use a “semi-continuous” (or “delta”,“hurdle” type) model:
  - Model the non-zeros as a gamma random variable
  - Zeros are modelled as a Bernoulli random variable
- For NPUE data, Poisson or negative binomial models can be used.

Statistical models

- Let w_i be O32 WPUE at station i. We define x_i and y_i as follows:

$$x_i = \begin{cases} 1 & w_i > 0 \\ 0 & w_i = 0 \end{cases}$$

$$y_i = \begin{cases} w_i & w_i > 0 \\ NA & w_i = 0 \end{cases}$$

$$x_i \sim \text{Bern}\left(p_i\right) \qquad y_i \sim \text{Gamma}\left(a_i, b_i\right)$$

which has mean $\mu_i = a_i / b_i$

Statistical models

- Let ε_i be a component of a Gaussian field (GF, a Gaussian process in two or more dimensions)
- ε_i connects the two processes as follows:

$$\operatorname{logit}(p_i) = \alpha_x + \varepsilon_i$$

$$\log(\mu_i) = \alpha_y + \beta_\varepsilon \varepsilon_i$$

- Here, α_x and α_y and are intercept terms and β_ε is a scaling parameter on the shared random effect.

Modelling dependence

- Spatial dependence is modelled through the ε_i

$$\boldsymbol{\varepsilon} \sim \mathrm{GF}\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

- That is, $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots \varepsilon_N]'$ is a Gaussian field with mean zero and covariance matrix $\boldsymbol{\Sigma}$.
- We use the flexible Matérn model for the spatial covariance model.
- Temporal dependence can also be modelled within INLA
  - We have used a simple AR(1) model in our modelling.

Including covariates

- Covariates are introduced into the zero and non-zero processes through the α_x and α_y terms.
- For example, we could consider linear relationships with station depth, D_i:

$$\alpha_x = \beta_0 + \beta_1 D_i$$

$$\alpha_y = \gamma_0 + \gamma_1 D_i$$

- At present, depth is the only covariate we have included in the modelling:
  - Simple parametric model not appropriate
  - We model relationships using a random walk model

Including covariates

- Other possible covariates include:
  - Survey date
  - Moon phase (tide strength)
  - Habitat information
  - Longitude and latitude
  - Oceanographic variables recorded on water column profilers, including pH, bottom temperature and dissolved oxygen
    - Work required to predict such oceanographic variables at stations where data are missing

Including covariates

- If goal is using WPUE and NPUE as density indices, important to consider which covariates affect density and which affect catchability.
- If a covariate affects catchability only, we predict at the same fixed value of that covariate for all stations.
- Many variables likely influence both density and catchability
  - This is one of the greatest challenges of modelling this type of catch-per-unit-effort data, when the purpose is to use the data in estimating indices of density or abundance.

Coverage gaps

- Areas that have had an expanded survey (2A, 4A in 2014, 4D edge in 2016)
  - Data allow us to estimate depth relationship over 10-400 ftm range
  - Model can predict WPUE in years when there is not full coverage
- Areas awaiting expansion (2B, 2C, 3A, 3B, 4B)
  - Predicting beyond the range of the data can lead to biased estimates
  - In areas adjacent to expanded areas (2B, 3B, 4B), deep/shallow data from neighbouring stations (in 2A, 4A) will inform the modelling
  - Is the best option is to continue the use of current coverage adjustment methods until an area has had a full survey expansion?

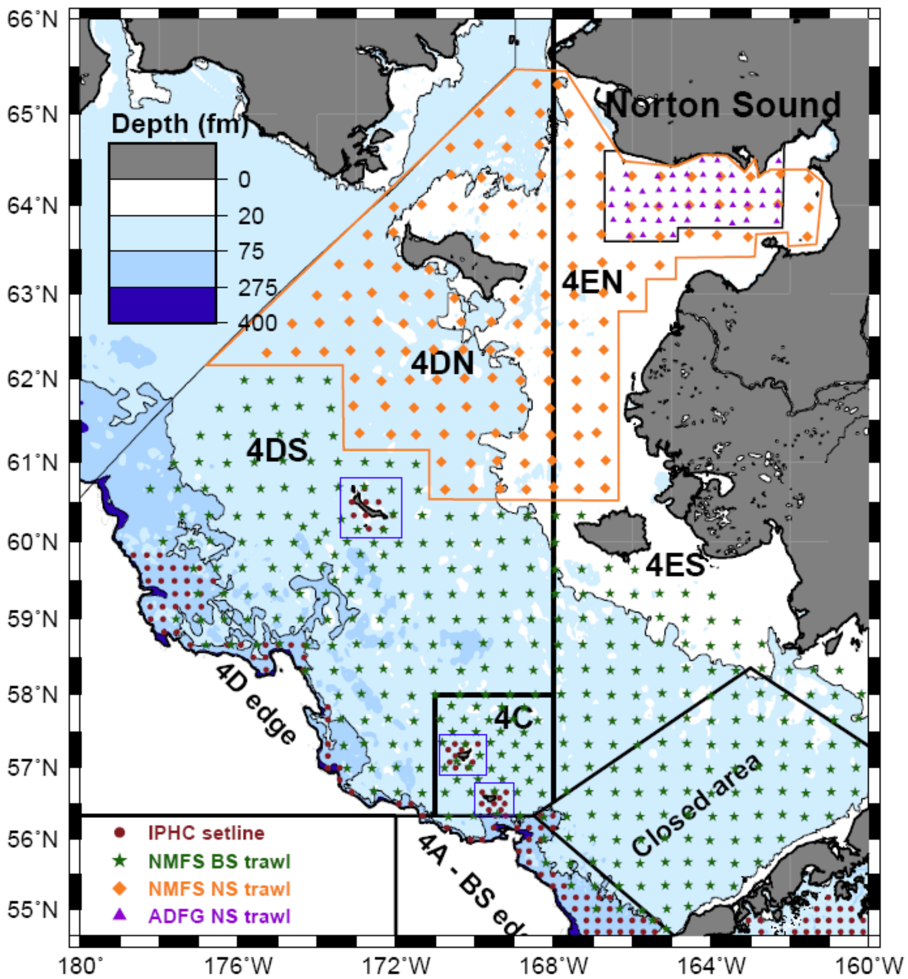
WPUE adjustment factors

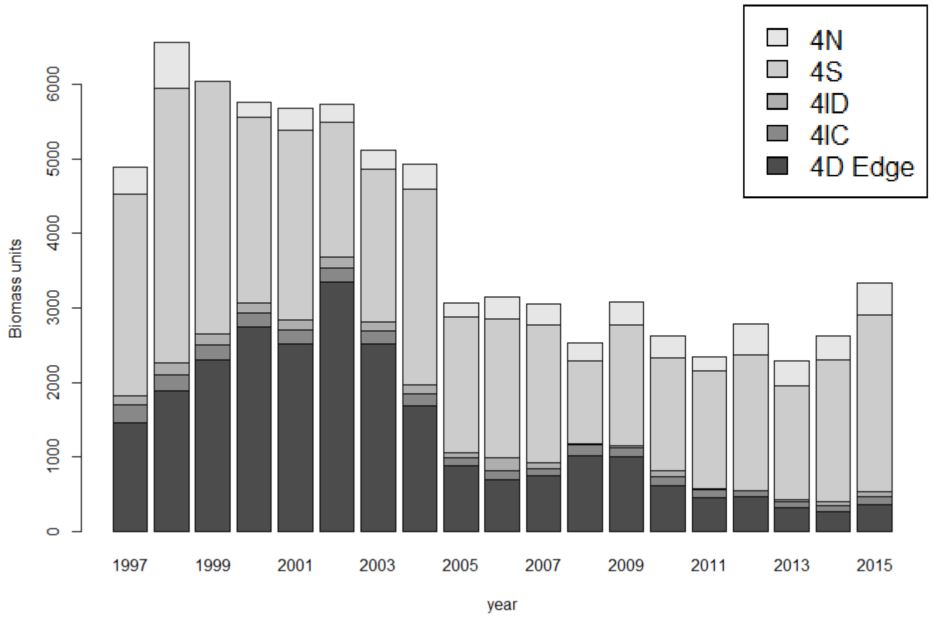
- Currently raw mean WPUE and NPUE are adjusted to account for:
  - Competition with halibut for baits
  - The timing of the survey relative to the harvest
- Data for the competition adjustment are available at the station level
  - Can use adjusted station WPUE and NPUE as data
  - Problem when zero baits returned
    - Adjustment is infinite
    - Use adjustment factor for 1 bait?
  - Need to compare results with current approach (adjustment factor computed from data aggregated over a regulatory area)

WPUE adjustment factors

- The timing adjustment includes data not available at fine spatial scales
  - E.g., sport harvest, personal use
- Still need to apply this adjustment to final regulatory area WPUE estimate when this is obtained as output from spatio-temporal modelling.

Special case: Area 4CDE





Special case: Area 4CDE

- Index for this area created from several component indices:
  - Area 4D Edge
    - will have full survey expansion data in 2016
  - Areas 4IC and 4ID
    - annual setline survey coverage since 2006
    - good spatial coverage within a small area
  - Area 4S (eastern Bering Sea flats)
    - sparse setline survey in 2006 and 2015
    - use calibrated NMFS trawl survey data in other years
  - Area 4N (northern Bering Sea)
    - use calibrated 2010 NMFS trawl survey data combined with triennial ADFG Norton Sound trawl survey

Area 4S

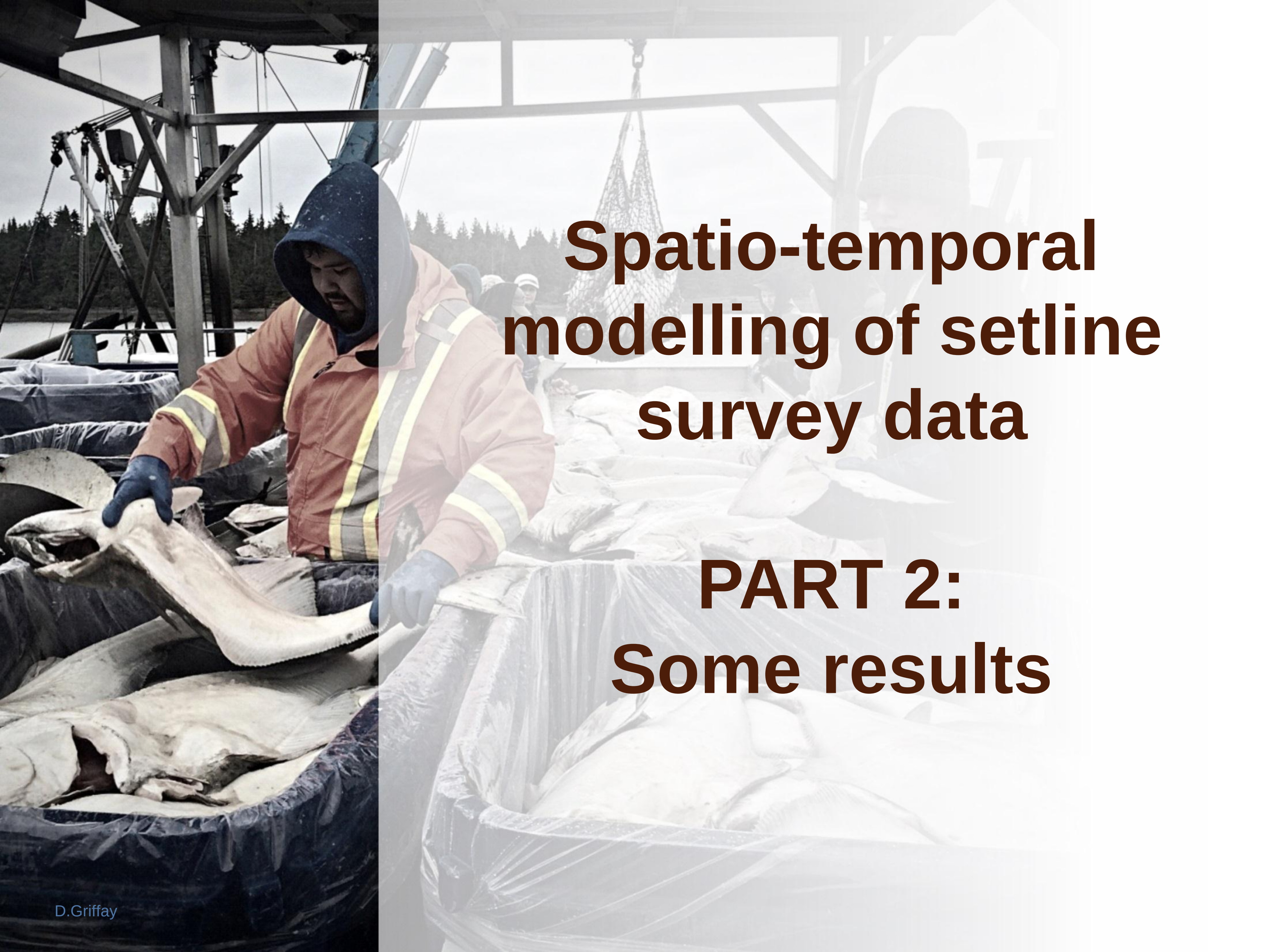
- Could use calibrated station-level trawl survey data in spatio-temporal modelling, along with setline survey data from 2006 and 2015.
- No simple way of estimating hook competition adjustment for calibrated trawl station data
  - In 2006 and 2015, apply same adjustment to all stations (setline and trawl)
  - In other years, we currently use the mean adjustment from 2006 and 2015, and this approach could be continued

Area 4N

- Can undertake spatial modelling of calibrated trawl data from 2010 NMFS trawl survey.
- Prediction of WPUE and NUPUE in other years will decrease in quality as time from 2010 increases.
- Norton Sound data do not consistently include individual halibut lengths necessary for application of the calibration
  - Assume all halibut have equal (non-zero) selectivity to include station data in spatial model

Spatio-temporal modelling of setline survey data

PART 2: Some results

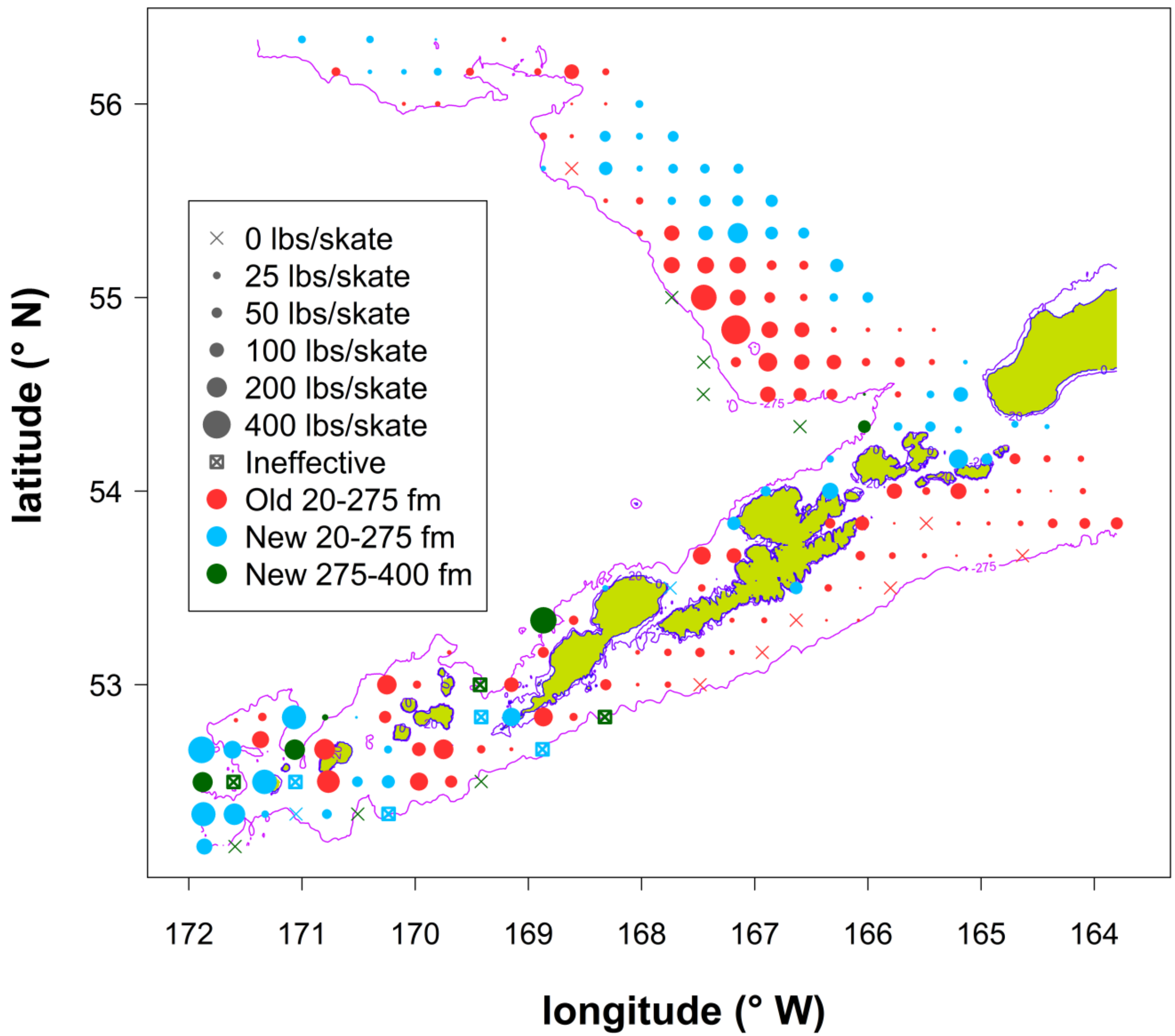


D.Griffay

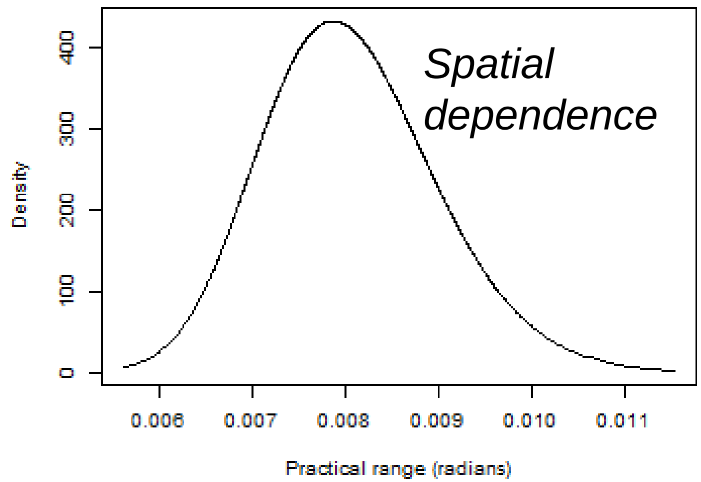
Estimating WPUE series

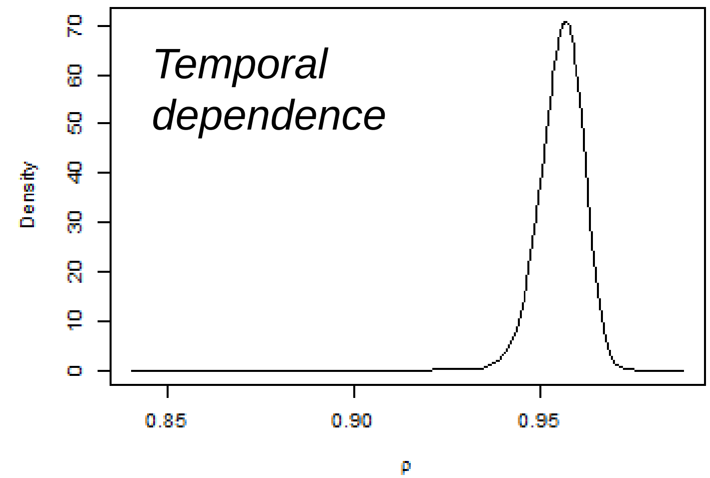
- At present, an estimated WPUE series is created by averaging model predictions at current and future survey station locations
  - Only potential stations in 10-400 ftm depths are used, consistent with the current survey expansion: no "potential" 0-10 ftm stations are used, although this is part of an area's bottom area
  - We are not currently integrating the predicted process over the entire bottom area
    - Could be done using a fine grid
    - Computationally intensive!

Case 1: Area 4A

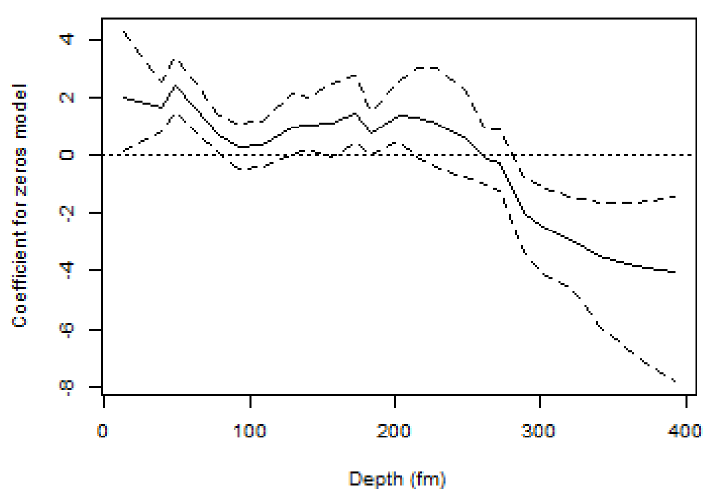


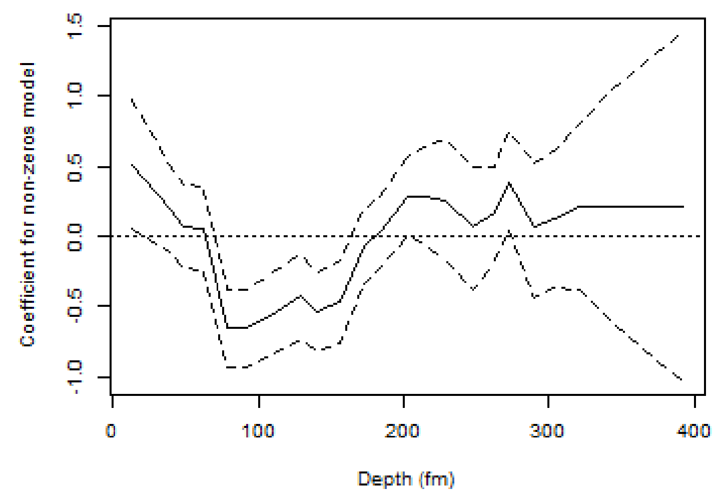
Case 1: Area 4A



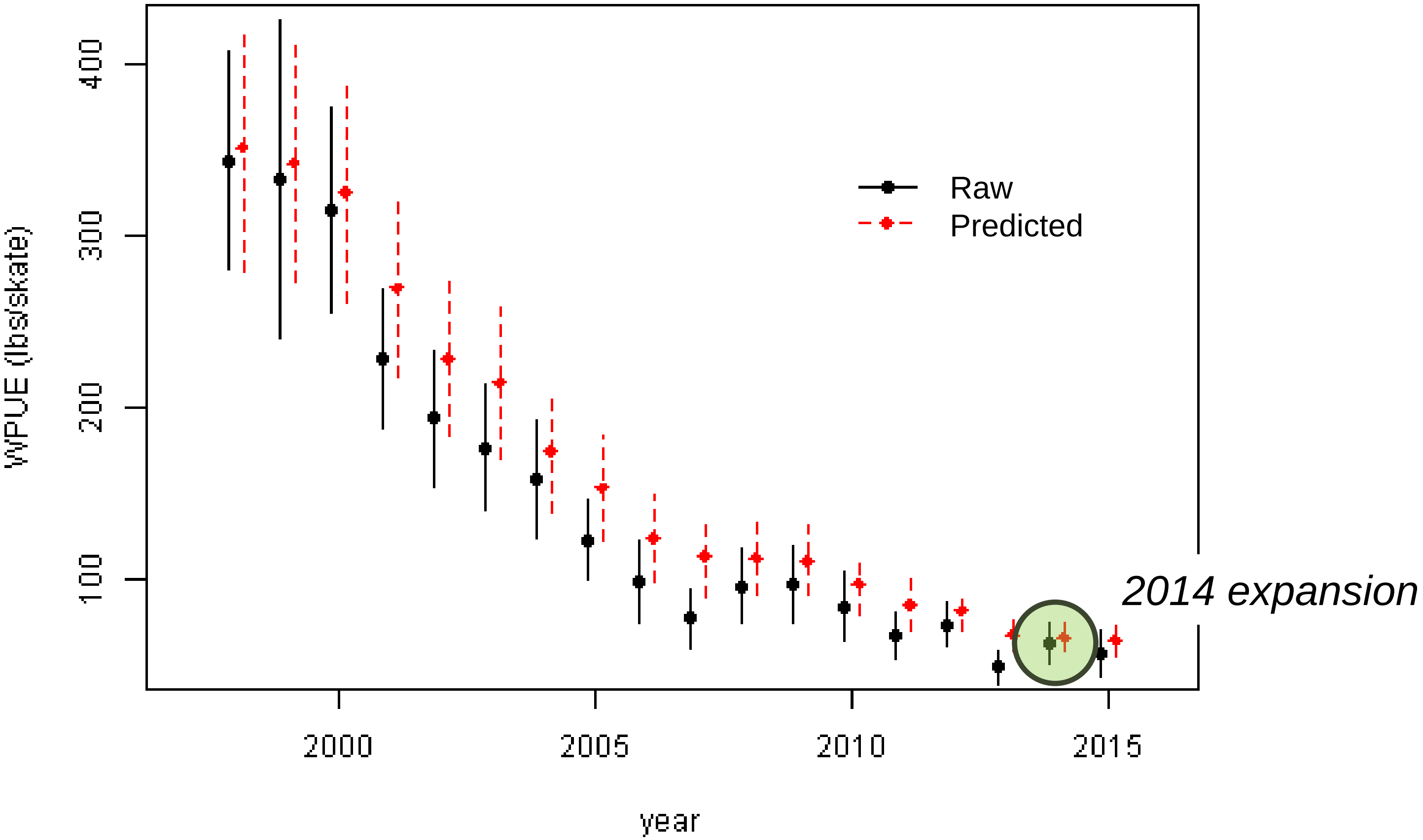


Station distance 10 nmi = 0.0029 radians



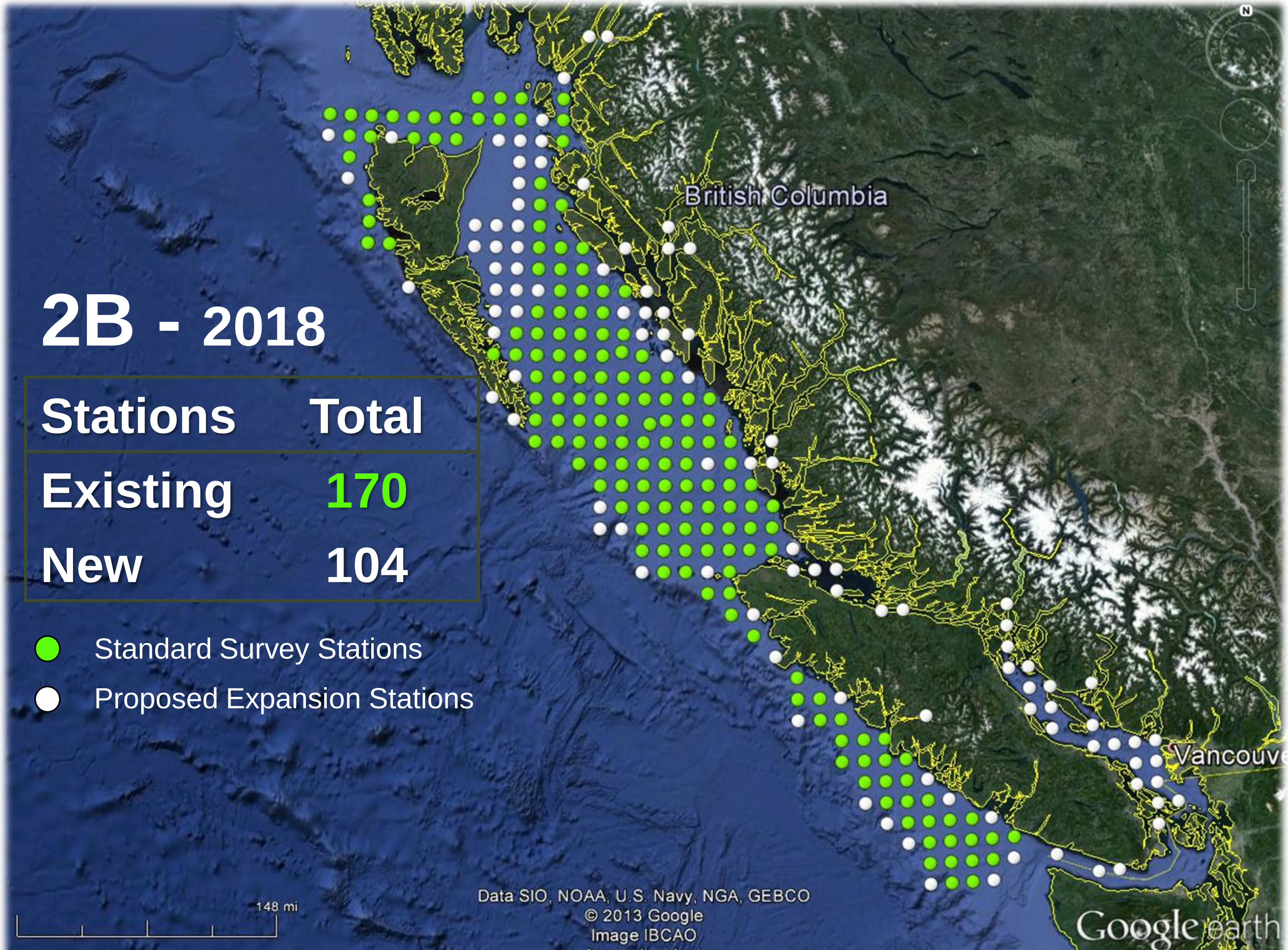


Case 1: Area 4A



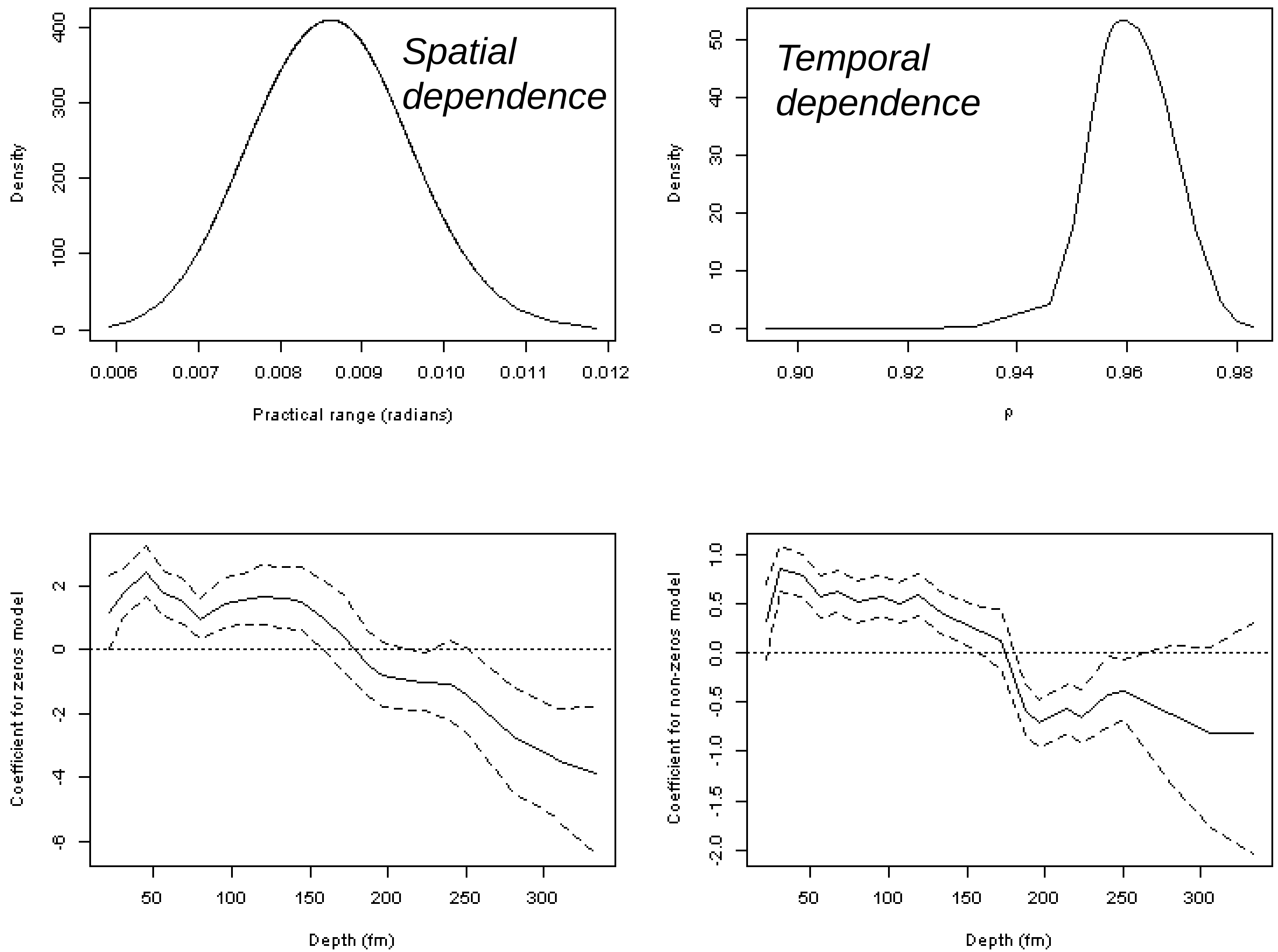
Switch to 4A pdf

Case 2: Area 2B

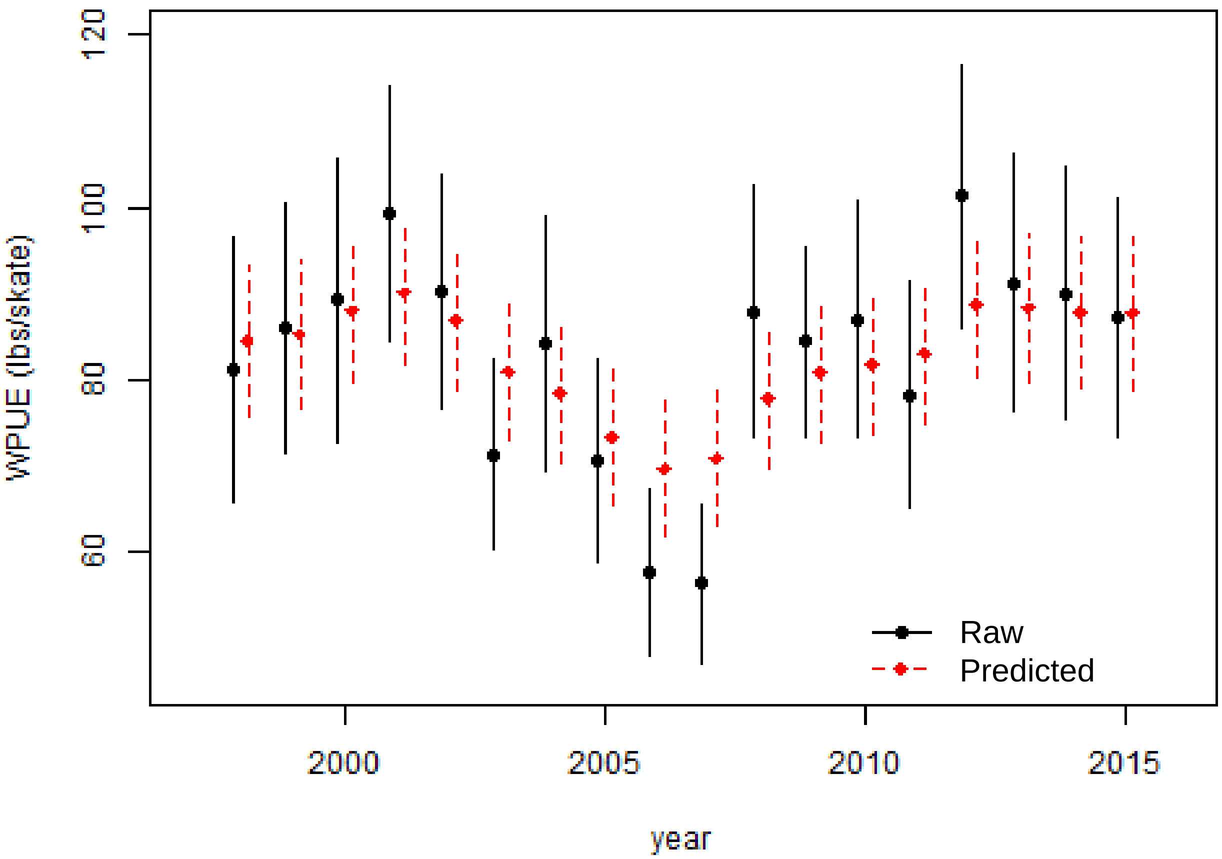


Stations	Total
Existing	170
New	104

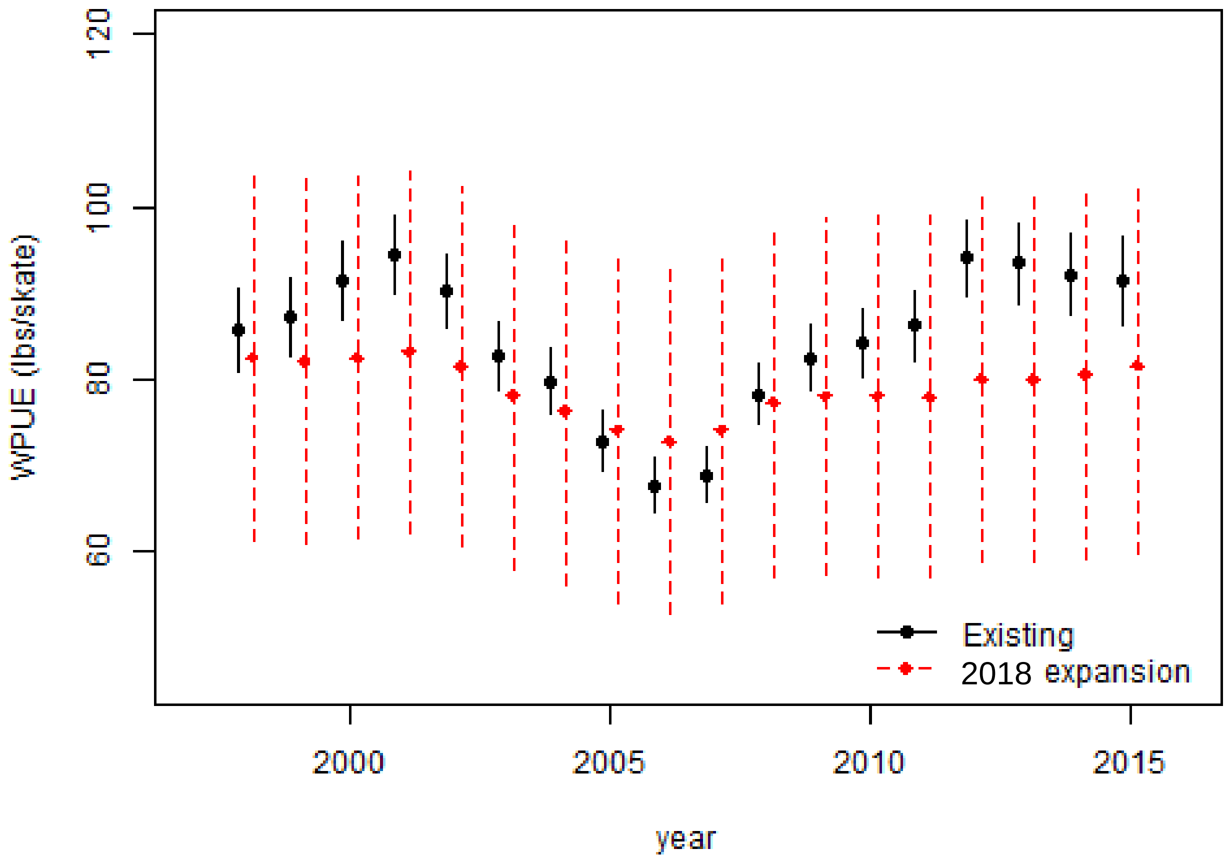
Example 2: Area 2B



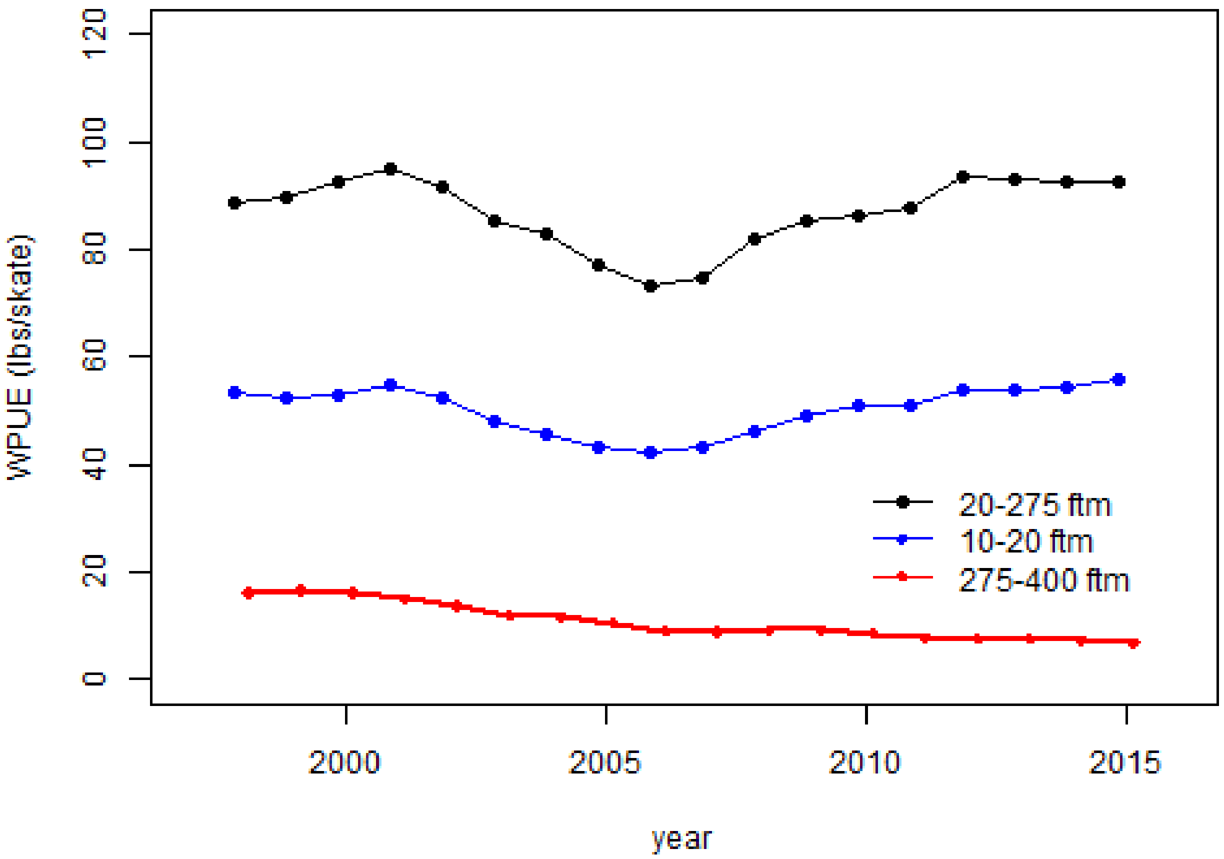
Case 2: Area 2B



Case 2: Area 2B



Case 2: Area 2B



Switch to 2B pdf

Does this modelling approach improve the density index?

- Accounts for uncertainty due to unsurveyed stations
  - Precision may or may not be improved over the current approach
  - However, we currently do not account for uncertainty in the various spatial coverage adjustment factors that are in use
- Reduces random variation in annual index values
  - We are modelling the underlying mean process, which has strong spatial and temporal dependence
- But prediction is poor at unsurveyed regions distant (in time and space) from surveyed stations
  - Low precision
  - Values are close to the process mean
  - *However*, the current approach is likely biased and ignores uncertainty in estimates in such regions

Conclusions

- The modelling approach offers clear advantages over the current method for estimating WPUE and NPUE from setline survey data.
- Along with the advantages listed on the last slide, we can dispense with the cumbersome and inelegant collection of adjustments for incomplete spatial coverage, leading to greater clarity and consistency in how these indices are calculated.
- We recommend use of the spatio-temporal modelling approach in 2016.

Discussion

- In areas yet to have an expansion do we…
  - Predict at unsurveyed locations?
  - Continue using adjustment factors?
- Do we compute the density indices using predictions at stations locations (past, present and future) or over the entire bottom area?

Discussion

- Hook competition adjustment
  - More work needed…
  - Should it be applied to the final series, or should we model adjusted station WPUE or NPUE?
- Timing adjustment
  - Do we continue using it?
  - Important for Area 2A only

Discussion

- Use of covariates
  - Which are most appropriate?
  - How can we avoid those that affect both density and catchability? (Or do they all do this to some extent?)

Discussion

- Separately model WPUE series, or use NPUE series times mean weight?
  - Station WPUE estimated using length-weight relationship
  - Nevertheless, using an area-wide mean weight ignores variation in mean size (length or weight) among stations
    - Does this matter? Variance has never been accounted for in the apportionment process!

Future work

- Modelling
  - Anisotropic models
  - Non-separable covariance structure
  - More covariates
  - Coastwide model (?!)
- Planning
  - Use real or simulated data to optimize design
  - How often should we fish the expansion stations?
  - Can we reduce the survey frequency in other regions?