**Methods Summary for Growth Subgroup Meeting**

**NOTE: This is excerpted from a larger manuscript draft.**

01 March 2019

We developed a data-driven analytical method to detect spatial changes in fish growth, tested it using simulated data, and then applied it to a commercially harvested groundfish stock in the Northeast Pacific. The method was designed to identify significant latitudinal break-points in the age-length relationship and not consider *a priori* hypotheses of spatial stratification.

The method involves fitting a Generalized Additive Model (GAM) using the mgcv package (Wood, 2011) in R (R Development Core Team, 2011) to the vector of observed lengths at age six as the response variable, predicted by separate smoothers for year and latitude., i.e.

Equation 1

where µt represents the expected mean of fish length, which is a random variable of which we have *t* observations; the linear effects of latitude () and year (), the last two of which are smooth functions. *g* is an invertible, monotonic function that enables mapping from the response scale to the scale of the linear predictor, such as the natural logarithm. To simplify the analysis, we fit the GAM to a subset of each simulated datasets including only female fish of age six (thus precluding the need to control for age or sex).

The first derivatives of the GAM with respect to latitude are evaluated to identify areas of significant change (i.e., break points) in growth parameter estimates. We used the method of finite differences (as in Simpson, 2018) to locate locations where fish size is changing the most. The finite differences approach approximates the first derivative of the spline generated from the GAM function. We then identify latitudes where the confidence interval of the first derivative is outside the 5th to 95th percentiles of the entire set of derivatives and do not include zero, rounded the value to the nearest integer, and designated these as “break points”.

## Simulation Testing

We performed a simulation study to evaluate the performance of the proposed method, using datasets generated using an individual-based model (IBM). The IBM (See Appendix for full details of IBM.) is capable of mimicking individual characteristics by following the life history processes (survival, growth, and reproduction) of individual fish. We simulate spatial variation by generating length and age datasets under different growth regimes (e.g. higher values of *K* and ) and assign a range of latitudes to each regime. The IBM implements the VBGF using Schnute’s formulation of the VGBF, which requires (*k*, *L*1, *L*2) so is computed as:

Equation 5 =

where represent the lengths of a fish at ages /, and *k* is the Brody growth parameter as before. An individual fish’s annual growth increment is subject to a bias-corrected lognormal error term. Depending on the scenario, the different regimes are either assigned completely distinct latitudinal ranges or ranges with some overlap. To simulate spatial zones, fish locations were sampled from a uniform distribution with boundaries specific to a certain growth ‘Regime’, or set of life history characteristics. Regime 1 refers to a central Pacific billfish-like species, where *L*∞ = 220cm and *k* = 0.258 in Regime 2 *L*∞= 350cm, Regime 3 *k* = 0.45, and Regime 3 *L*∞= 250cm, *k* = 0.3. In all except the final scenario, the latitude of fish grown under Regime 1 sampled from a uniform distribution between 0° and 25°; For Regimes 2 and 3 latitude is sampled uniformly from 25° to 50°. In the final scenario, fish simulated under life history Regime 1 are assigned latitudes sampled uniformly from 1° to 49°, and those simulated under Regime 2 have locations uniformly sampled from 49° to 50°.

The simulation scenarios described in Table 1 were designed to represent a variety of possible regimes in spatial growth variation. Under each scenario, we generated 100 replicate datasets which averaged 720 age-six fish per dataset, and tabulated the frequency at which a given (true) break point was identified using the method described above. The method was evaluated based on: a) if it was able to accurately detect the presence or absence and location of ‘break point(s)’ in space, and b) the coverage probability of VGBF parameter estimates when data were re-aggregated at the proposed break point. Figure 1 illustrates the latitudinal distribution of age-six fish for an example of one simulated dataset (of 100) under each scenario.

|  |  |
| --- | --- |
| **Scenario Description** | **Spatial Stratification** |
| No spatial breaks | Latitude is uniform random variable for all observations |
| Single, spatial break in middle of range, with no overlap and strong contrast | Latitude ~ U[0,25] under growth Regime 1; Latitude ~ U[25,50] under Regime 2 |
| Single, spatial break at 25 degrees with no overlap and reduced contrast | Latitude ~ U[0,25] under growth Regime 1; Latitude ~ U[25,50] under Regime 3 |
| Single spatial break with some overlap | Latitude ~ U[0,25] under growth Regime 1; Latitude ~ U[20,50] under Regime 2 |
| Single spatial break at edge of range with no overlap | Latitude ~ U[0,49] under growth Regime 1; Latitude ~ U[49,50] under Regime 2 |

Table 1. Summary of simulated datasets used to test the proposal method in presence/absence of spatial variation in growth. Regime 1 refers to a central Pacific billfish-like species, where *L*∞ = 220cm and *k* = 0.258; In Regime 2 *L*∞= 350cm, *k* = 0.45, and Regime 3 *L*∞= 250cm, *k* = 0.3.

For each simulated dataset, we constructed a GAM function as described above, utilized the method of finite differences to identify time periods and/or locations where growth was changing quickly (where the confidence interval of the first derivative was outside of the 5th to 95th percentiles of the entire derivative set for that simulation, **G** and did not include zero). The data were then re-aggregated to match the most frequently identified spatial break point for that dataset, and fit to a VGBF in TMB to provide estimated life history parameters L∞ and k. The ‘coverage probability’ was computed as the proportion where the estimated 95% confidence interval for L∞ and k contained the true value

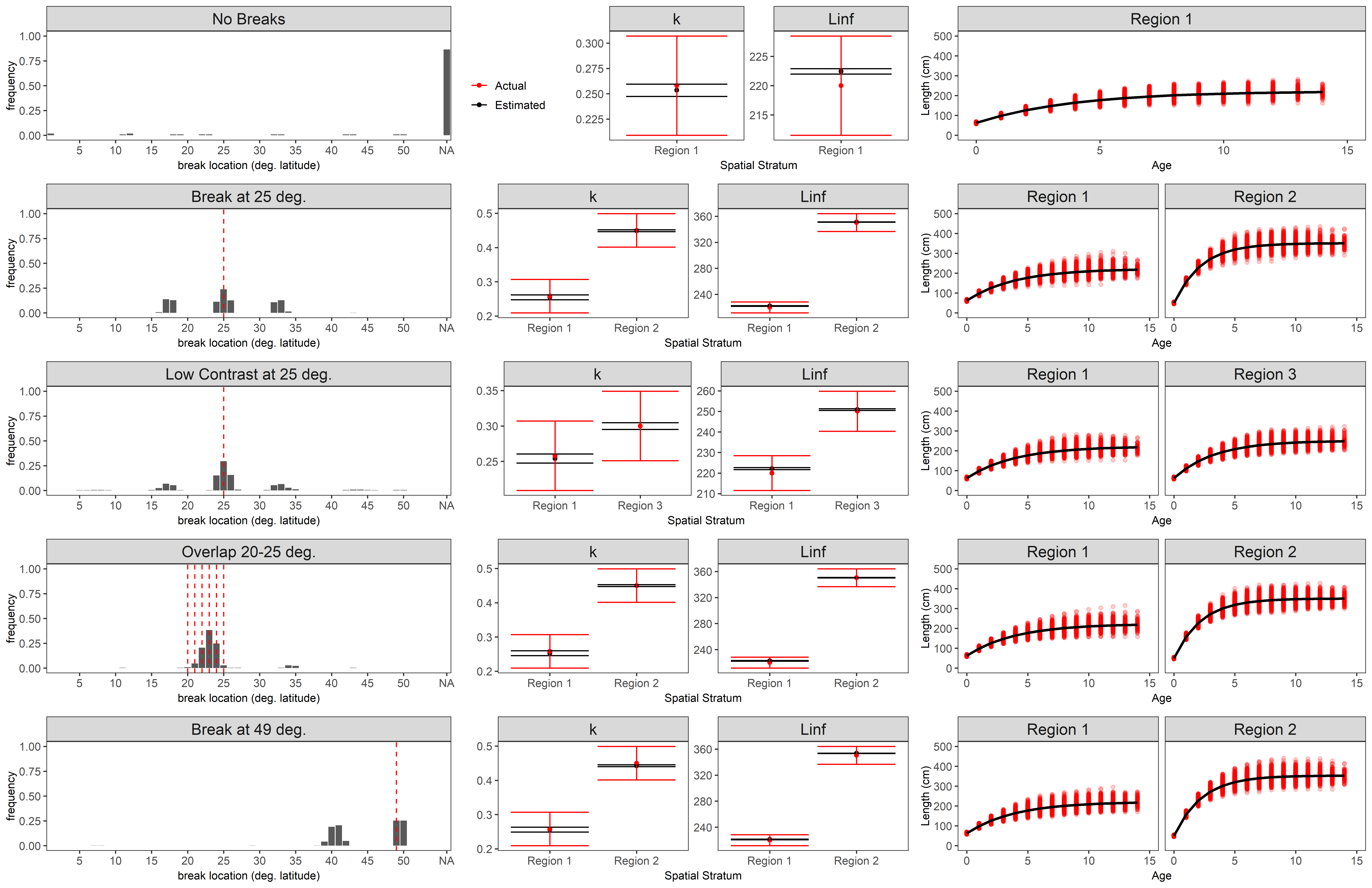


Figure 1 Results of simulation analysis. Each row corresponds to a tested scenario; left-hand panels show a frequency histogram of latitudinal break points identified using the GAM methods, with vertical dashed red lines indicating the underlying true break point. A break point of “NA” indicates that no break point was detected. Center panels (second columns) compare the original and fitted life history parameters (*L*∞ and *k*) used to generate the data and parameters estimated when the data were aggregated at the most commonly observed break (black points). Error bars represent 95% confidence intervals. The rightmost panel displays the original data (red points) and fitted VBGF (black lines).

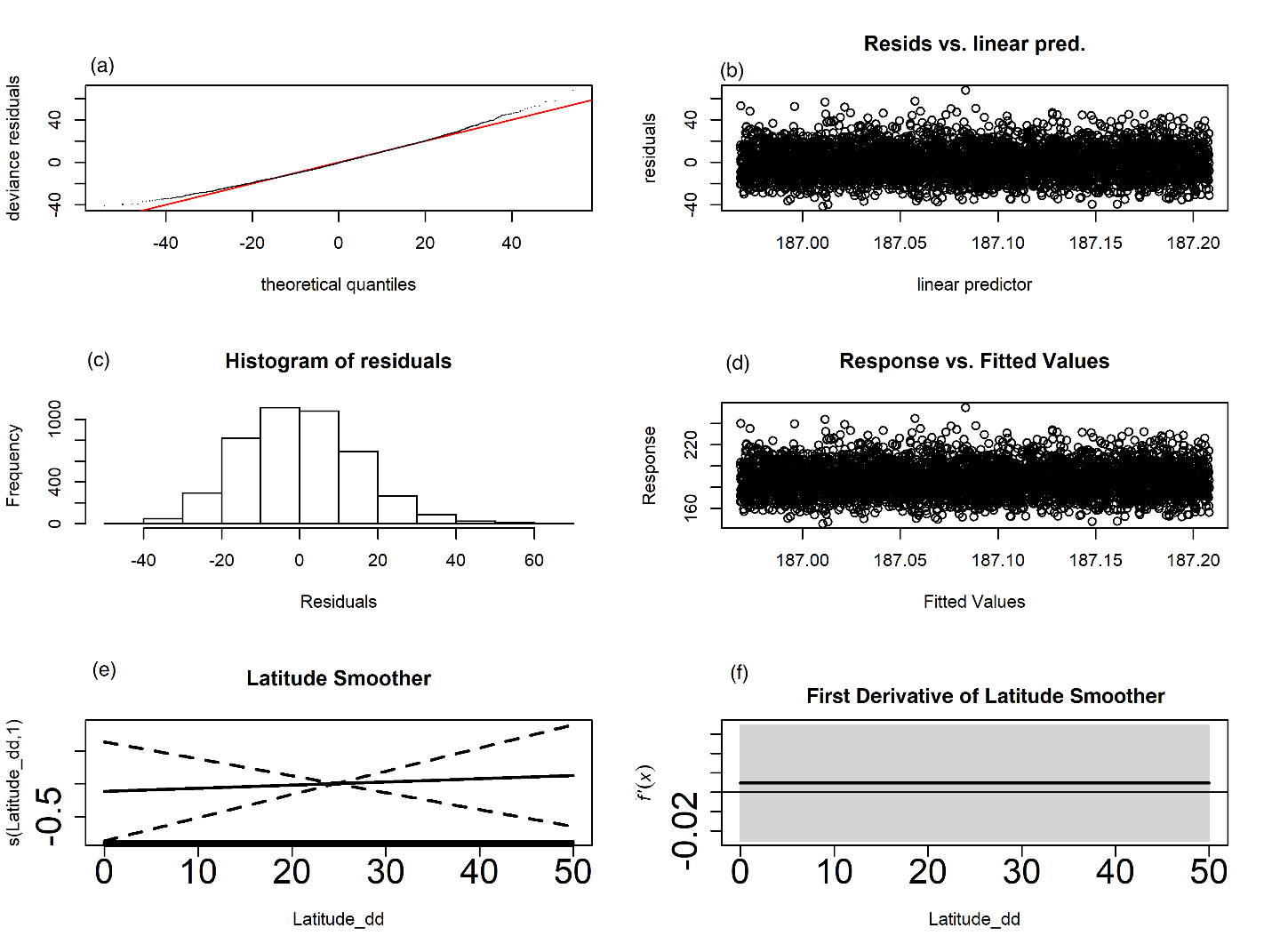


Figure 1. (a-d) Diagnostics of GAM, (e) the raw value of the latitudinal smoother, and (f) the first derivative of the latitudinal smoother for a single simulated dataset with no designated breaks. The grey shaded region in (f)indicates the 95% confidence interval for the first derivative.

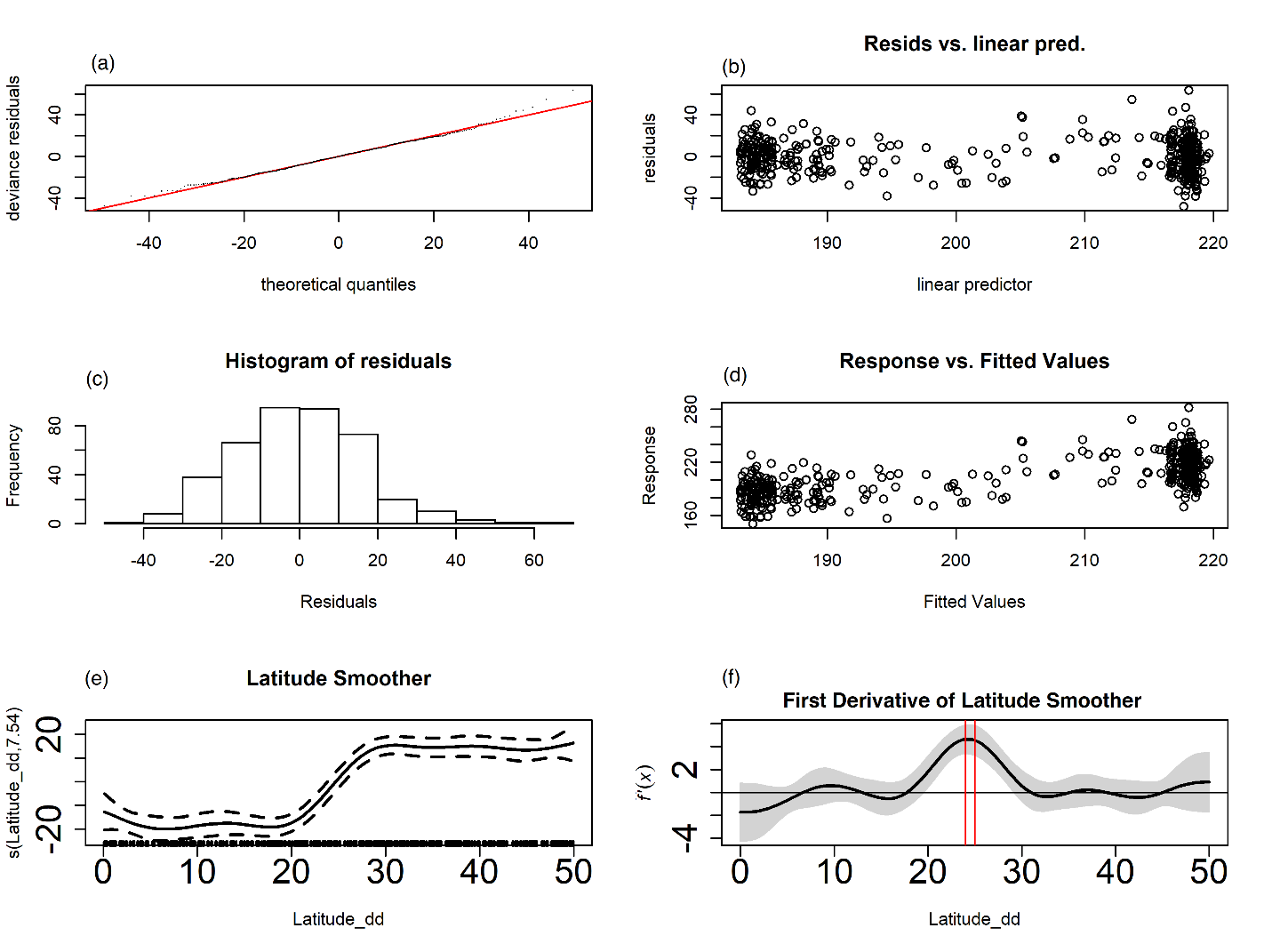


Figure 2. (a-d) Diagnostics of GAM, (e) the raw value of the latitudinal smoother, and (f) the first derivative of the latitudinal smoother for a single simulated dataset with a low-contrast break point at 25 degrees. The grey shaded region in (f) indicates the 95% confidence interval for the first derivative; vertical red lines indicate detected break points, which are outside the 95th percentile for this dataset and do not have a confidence interval that contains zero.

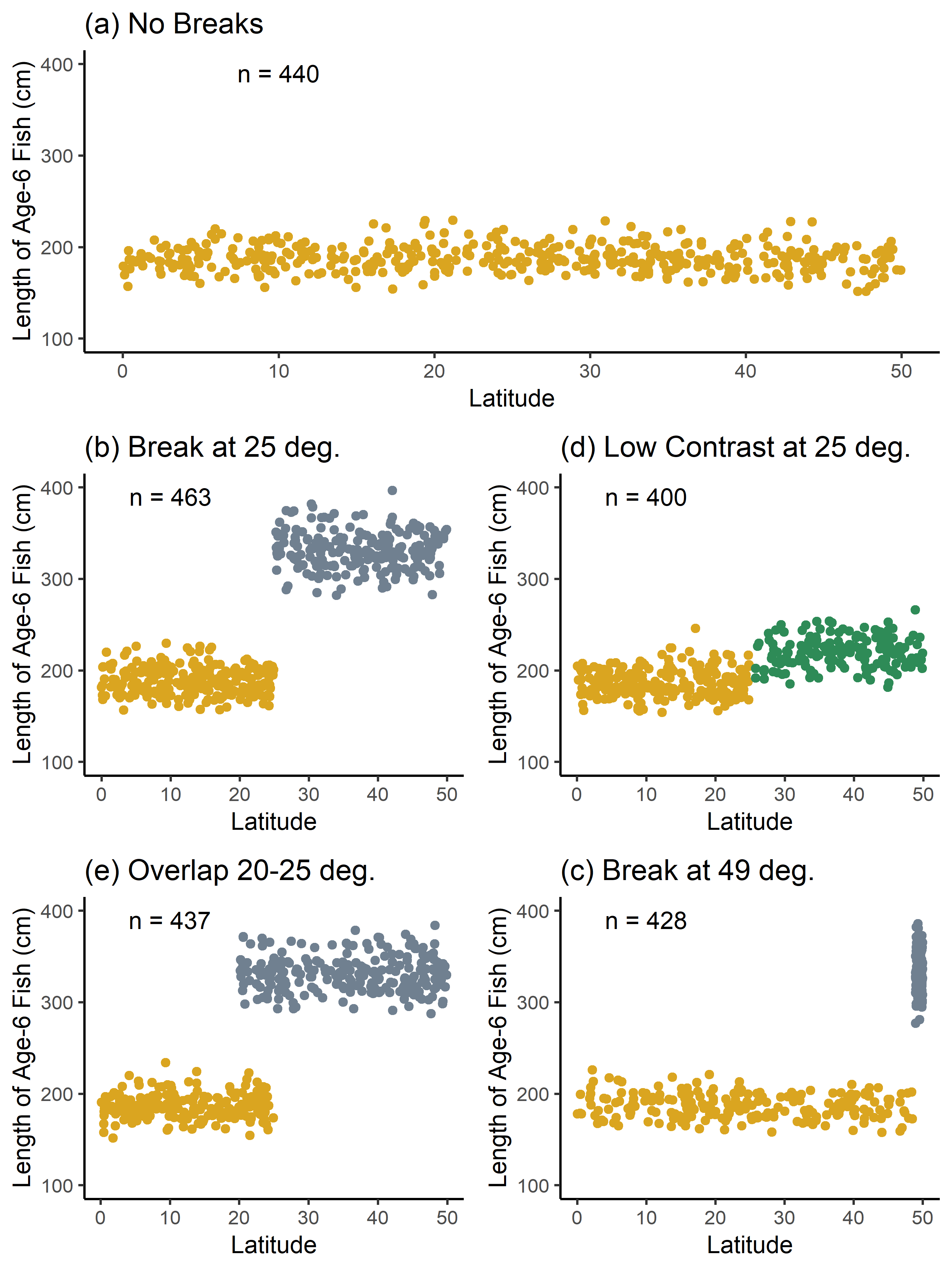


Figure 3. Example single dataset for each tested spatial scenario presented in Table 1. For each scenario, points represent the length and latitudinal location of a single simulated fish of age six. Fish locations (latitudes) were sampled from a uniform distribution of the boundaries indicated in Table 1. Text labels indicate the number of individual fish in the sample.