Methods for spatial survey modelling - program of work for 2019

PREPARED BY: IPHC SECRETARIAT (R. WEBSTER; 24 AUGUST 2019)

PURPOSE

To address Scientific Review Board (SRB) requests regarding the proposed methods for assessing options for a rationalised IPHC fishery-independent setline survey (FISS).

INTRODUCTION

The IPHC has been undertaking a series of fishery-independent setline survey (FISS) expansions, beginning with a 2011 pilot in IPHC Regulatory Area 2A, and continuing from 2014-19 as follows:

- 2014: Regulatory Areas 2A and 4A
- 2015: Regulatory Area 4CDE eastern Bering Sea flats
- 2016: Regulatory Area 4CDE shelf edge
- 2017: Regulatory Areas 2A and 4B
- 2018: Regulatory Areas 2B and 2C
- 2019: Regulatory Areas 3A and 3B

The purpose of the expansion program has been to fill in the often large gaps in the annually-fished FISS to build a complete picture of Pacific halibut density throughout its range, and thereby reduce bias and improve precision in density indices and other quantities computed from the FISS data.

With the planned expansions due for completion in 2019, the intention is to use our improved understanding of the Pacific halibut distribution to re-design the annual FISS. As a result, it is likely that stations that were previously fished annually may require less frequent fishing, while it may be preferable to annually fish some expansion stations that have been surveyed just once to date.

In report IPHC-2019-SRB014-05, we proposed criteria and methods for evaluating such a FISS rationalisation. In response, the SRB had the following request:

> *"The SRB requested analysis of past prediction patterns (a type of cross-validation analysis) to help assess the proposed methods' ability to meet precision targets while maintaining low bias. This should include an examination of spatio-temporal residual patterns for the appropriateness of estimated autocorrelation."* (IPHC-2019-SRB014-R).

This report addresses the SRB request.

Methods

Past prediction patterns

The overall goal of the FISS rationalisation is to maintain or enhance data quality (precision and bias) subject to the cost constraints of the FISS budget. In IPHC-2019-SRB014-05, we proposed targets on the coefficient of variation by management unit (IPHC Regulatory Area, Biological Region, coastwide stock) to ensure data quality in a rationalised FISS design. We proposed evaluating potential future designs by fitting models to simulated data sets created from samples from the posterior predictive distributions from the most recent year's spatio-temporal modelling output. The CVs of the estimated mean weight or numbers per unit effort indices (WPUE or NPUE) could be estimated from models fitted to the observed data augmented with simulated data for future years to assess whether proposed designs were likely to meet precision targets.

To better understand whether the CVs estimated from data sets augmented with simulated data provide a useful assessment of likely future data quality, we re-fitted models to the 1993-2018 data series with the 2018 observations replaced by simulated data derived from the 2017 posterior predictive distributions. We did this for both IPHC Regulatory Areas 2A and 4B, which were used as examples in the presentation at SRB014. Because we can expect some between-simulation variation, we fitted models using three different simulated data sets to give some sense of consistency.

Posterior predictive diagnostics

The request to examine spatio-temporal residual patterns falls within the area of posterior predictive diagnostics, in which discrepancy measures are used to assess model fit (Cressie and Wikle, 2011). Using the notation of Cressie and Wikle (2011), we make use of the discrepancy measure, *T*:

$$T(Z_i;Y,\theta) = \frac{\left(Z_i - E(Z_i \mid Y,\theta)\right)^2}{\mathrm{var}(Z_i \mid Y,\theta)}$$

Here Z_i is observation *i* (i.e., observed WPUE or NPUE), *Y* is the underlying process, and θ is the vector of model parameters. $E(Z_i \mid Y,\theta)$ and $\mathrm{var}(Z_i \mid Y,\theta)$ are the mean and variance respectively of the posterior predictive distribution of Z_i. The value of *T* for the observed Z_i can be compared to the values computed by substituting the samples from the corresponding posterior distribution (we save 2000 such samples as part of the model output). As a measure of "extremeness" we compute the posterior predictive *p*-value for each observation Z_i,

$$P\left(T(Z_{i,rep};Y,\theta) \geq T(Z_i;Y,\theta) \mid Z_i\right)$$

where the $Z_{i,rep}$ are the posterior samples. Very small *p*-values are indicative of local lack of fit. In this report, we spatially visualise the discrepancy values by mapping *T* for each observed station over several years for two example IPHC Regulatory Areas (4B and 4CDE).

Results

Table 1 compares the estimated CVs for O32 WPUE from the 2018 modelling of the 1993-2018 data series, to those estimated by replacing the observed 2018 data with simulated data sets created from the 2017 posterior predictive distributions. One decimal place is used in these comparisons to get a more accurate measure of the size of any differences as they are relatively small.

Table 1. CVs (%) for 2018 O32 WPUE estimated using the full 1993-2018 data series, and using the 1993-2017 series augmented with simulated data for 2018.

Regulatory Area	1993-2018 data	1993-2017 data, 2018 simulated		
		Simulation 1	Simulation 2	Simulation 3
2A	11.7%	10.8%	10.3%	11.0%
4B	14.1%	12.9%	13.4%	13.8%

For IPHC Regulatory Area 4B from 2011-18, Figures 1-3 map the discrepancy values, T, with plot symbol area proportional to T, and with purple symbols indicating the most extreme values (posterior predictive $p<=0.05$). While each year shows evidence for local lack of fit at several locations, there do not appear to be consistent spatial patterns among years. Figures 4-11 plot values of T for IPHC Regulatory Area 4CDE, with colours again highlighting those with very low p-values.

Discussion

While the CVs estimated for 2018 using simulated data are consistently lower than those estimated using observed data, the differences are not great enough to suggest that the use of posterior samples to predict precision will meaningfully affect the comparison of future design options.

The maps of discrepancy values for the example Regulatory Areas of 4B and 4CDE show evidence for localised lack of fit in the space-time model. While there are no consistent patches of poor fit apparent in the IPHC Regulatory Area 4B maps, for Regulatory Area 4CDE, parts of the Bering Sea shelf edge in the west, and around Pribilof and St Matthew Islands contain clusters of stations with high discrepancy values and low posterior predictive p-values in most years. The fitted space-time models at present feature parameters that characterise spatial dependency that do not vary in time and space. Thus, we are essentially estimating an average for these parameters across each IPHC Regulatory Area, when (at least for 4CDE), the strength of spatial dependence may vary somewhat across space. In IPHC Regulatory Area 4CDE, it seems likely that spatial dependence is weaker in the relatively high-density regions along the shelf edge and Bering Sea islands than it is in the more uniform habitat of the shallow waters that comprise the greater part of the eastern Bering Sea. More flexible models that allow the strength of spatial dependence to vary within a Regulatory Area are under consideration for future modelling work, particularly in the Bering Sea.

Otherwise, the apparent randomly located stations with high values of T may indicate that the current process model within the space-time models may be understating the probability of observing very high (or low) WPUE at a station relative to nearby stations. Observations are assumed to come from a mixture of Bernoulli (for the probability of non-zero WPUE) and gamma (for the distribution of non-zero WPUE) distributions. It is possible the probability of very high (or low) WPUE is greater than expected under the gamma model. Future work should explore alternatives for this component of the space-time models used for survey data.

Finally, the R-INLA package used for model implementation comes with options for efficient leave-one-out cross validation methods for assessing model fit, which may be preferable to the posterior predictive checks used above (Held et al., 2009). These have yet to be explored with our space-time models, but will be added to the model runs for 2019.

References

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-temporal Data*. John Wiley and Sons, Hoboken, New Jersey.

Held, L., Schrodle, B, and Rue, H. (2009). Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. In Kneib, T. and Tutz, G. (eds) *Statistical Modelling and Regression Structures: Festschrift in honour of Ludwig Fahrmeir.* Physica-Verlag HD.

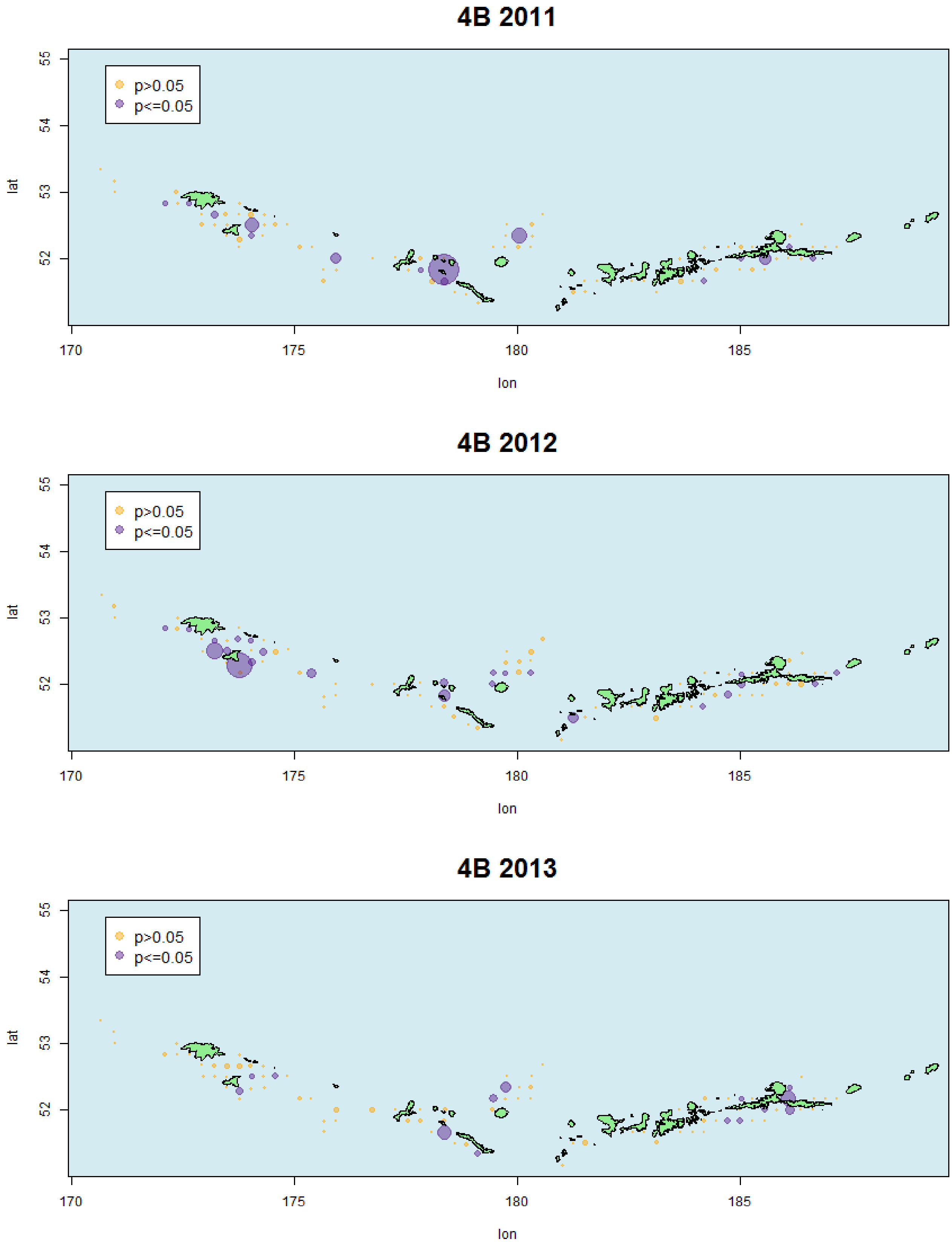


Figure 1 Maps of FISS stations with symbols showing relative sizes of discrepancy measures, *T*, and colours distinguishing those with very low *p*-values ($p<=0.05$), for Regulatory Area 4B, 2011-13.

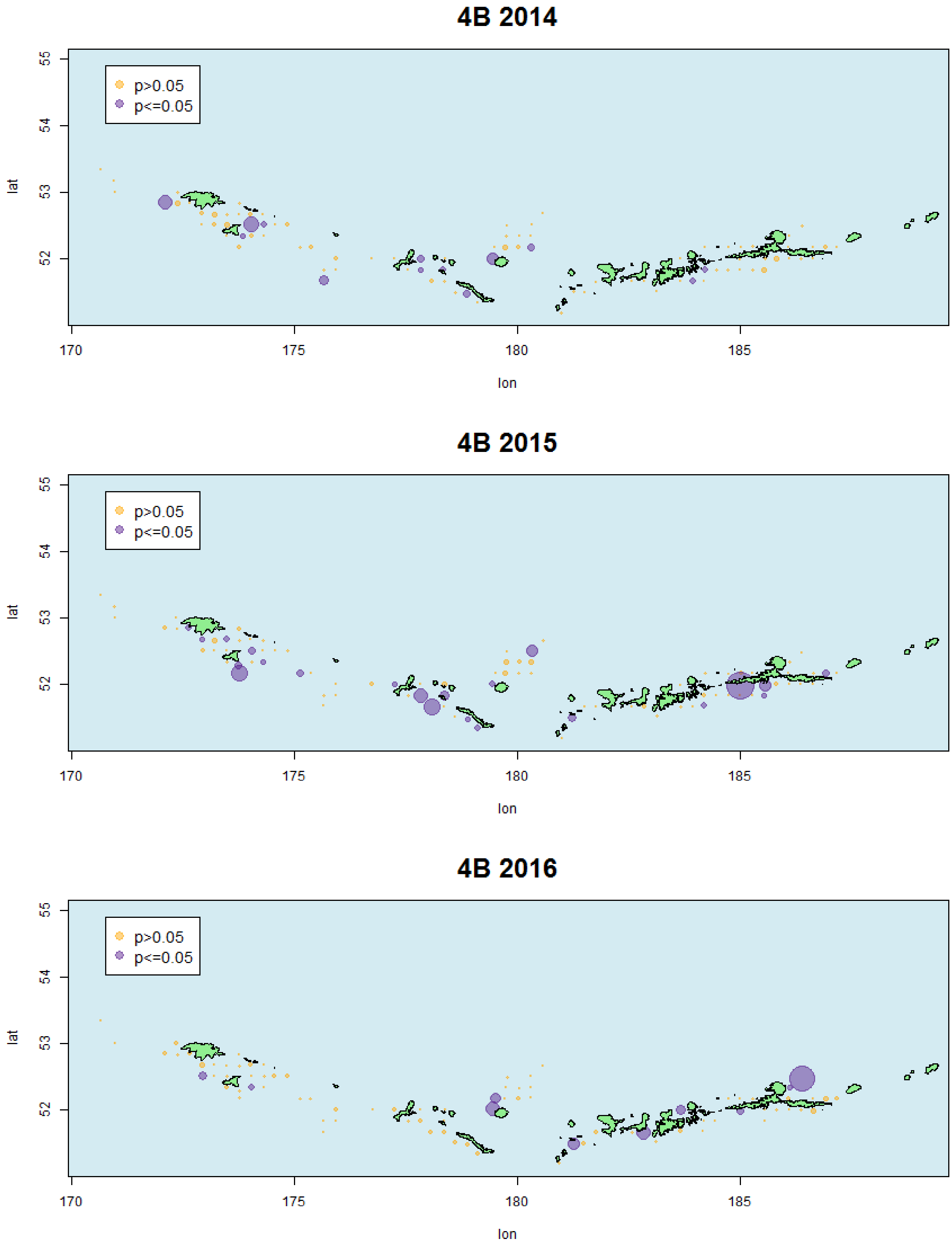
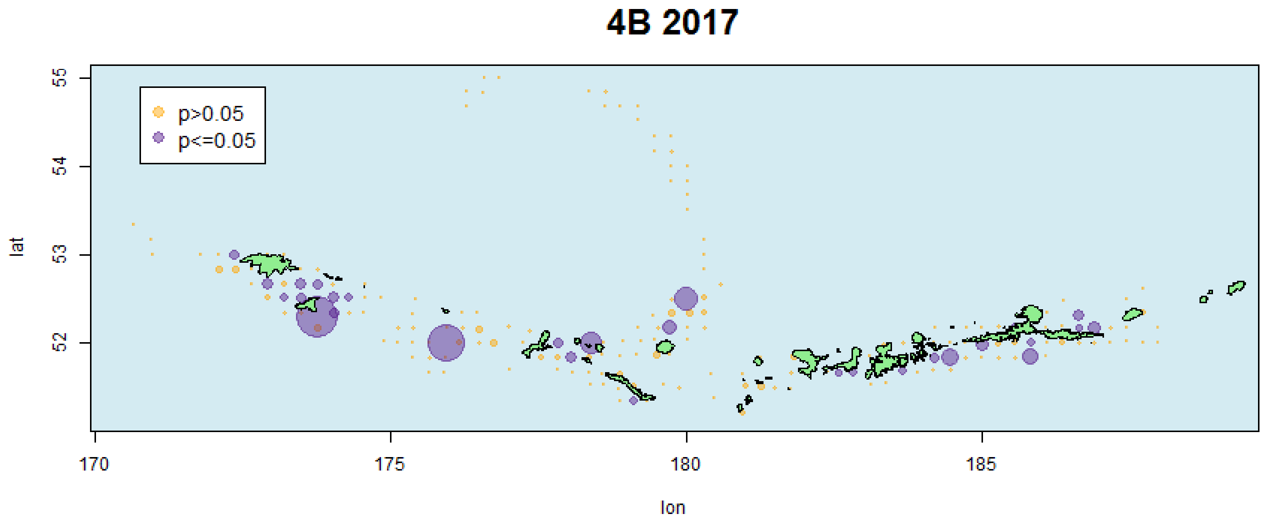


Figure 2. Maps of FISS stations with symbols showing relative sizes of discrepancy measures, *T*, and colours distinguishing those with very low *p*-values ($p<=0.05$), for Regulatory Area 4B, 2014-16.



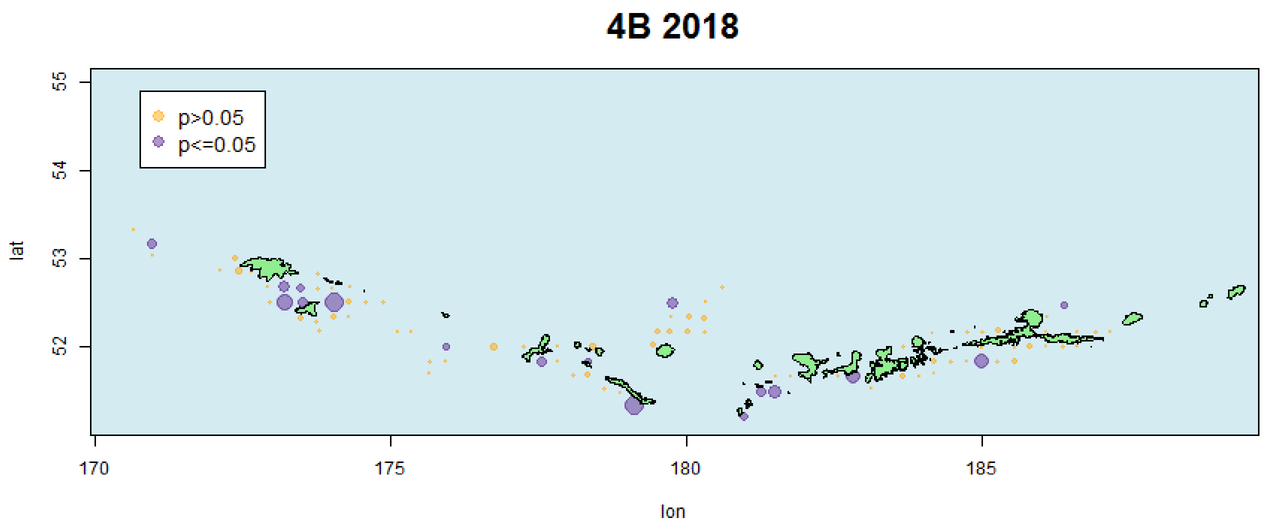


Figure 3. Maps of FISS stations with symbols showing relative sizes of discrepancy measures, *T*, and colours distinguishing those with very low *p*-values (p<=0.05), for Regulatory Area 4B, 2017-18.

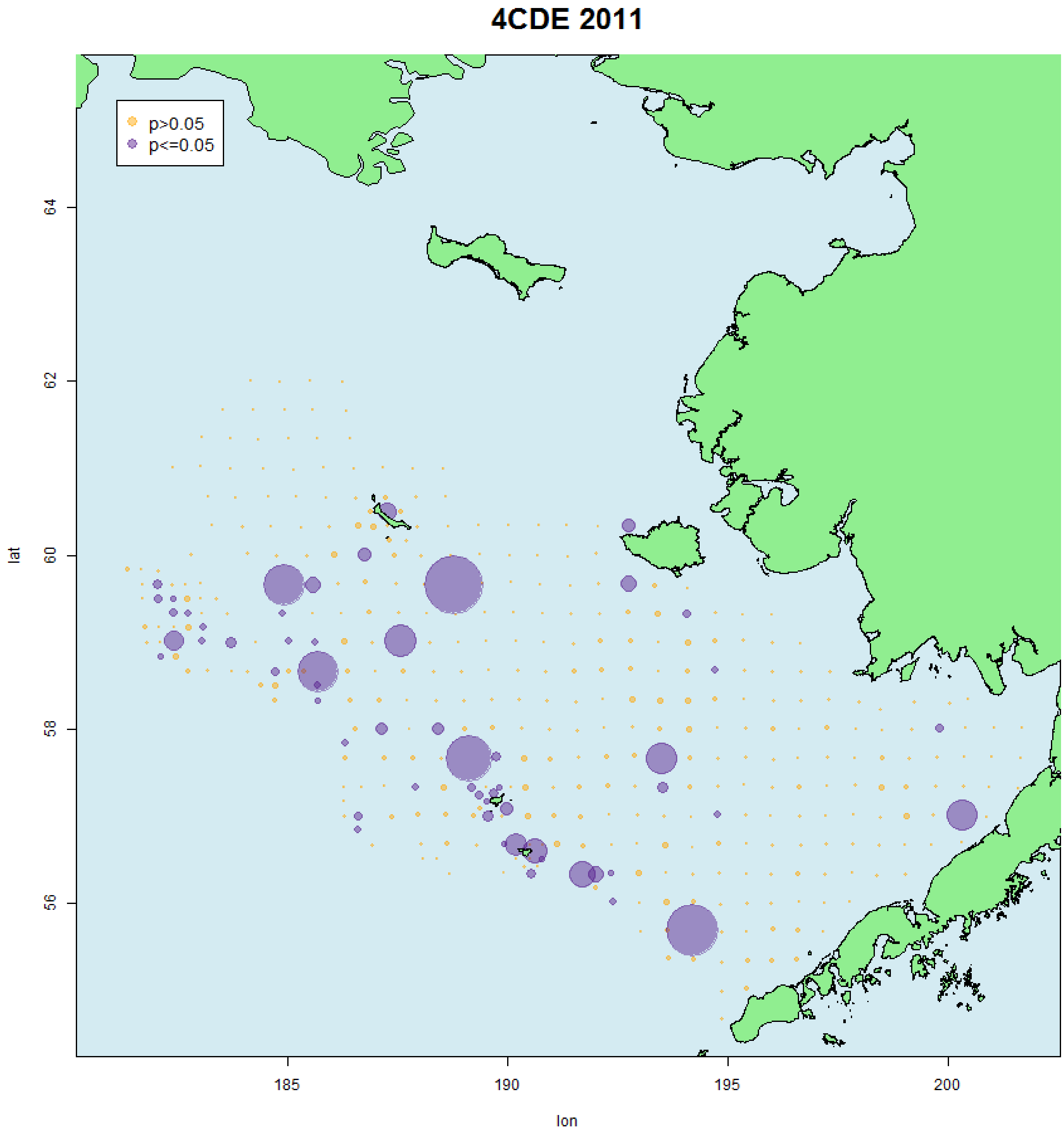


Figure 4. Maps of FISS stations with symbols showing relative sizes of discrepancy measures, *T*, and colours distinguishing those with very low *p*-values (p<=0.05), for Regulatory Area 4CDE, 2011.

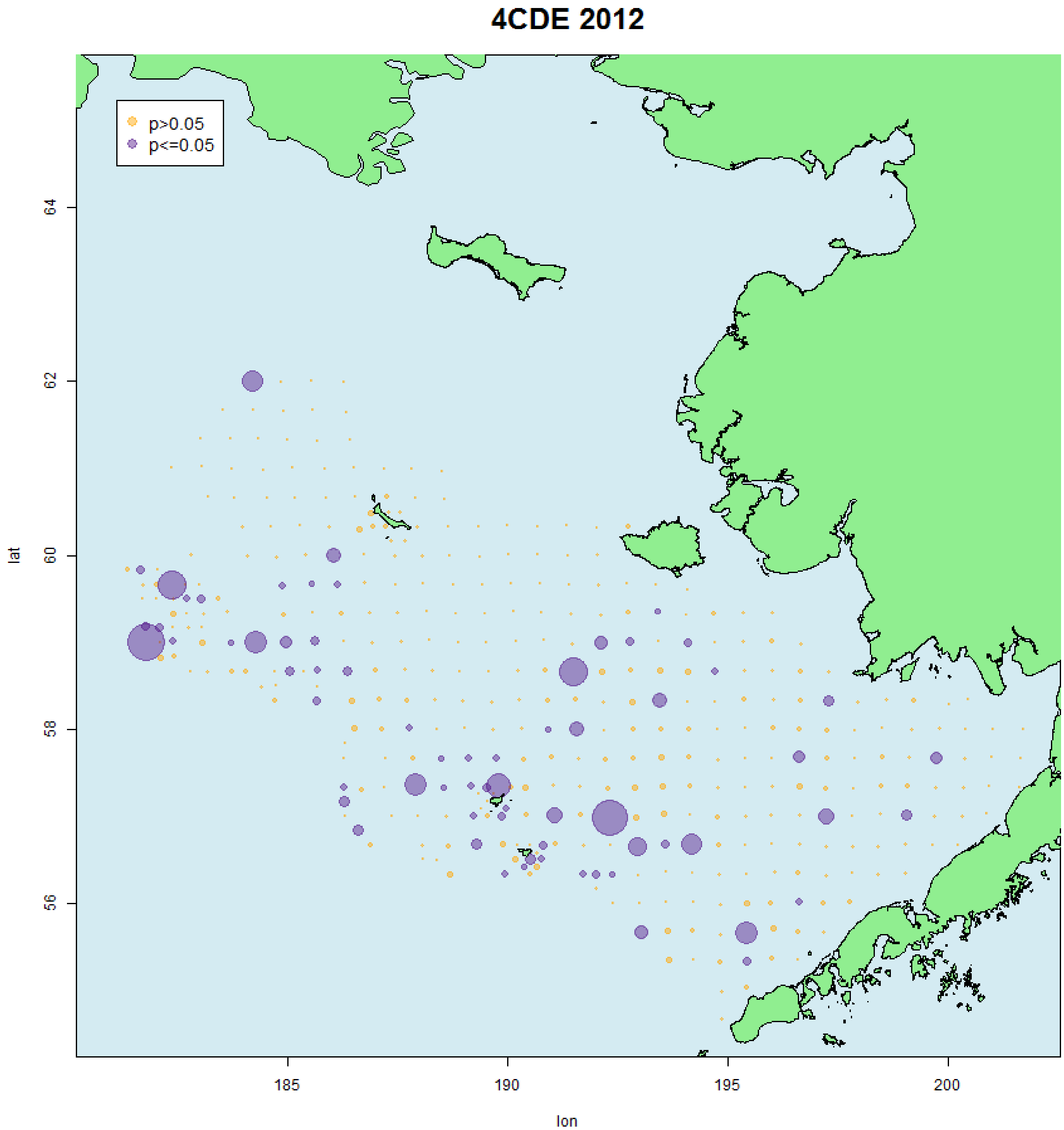


Figure 5. Maps of FISS stations with symbols showing relative sizes of discrepancy measures, *T*, and colours distinguishing those with very low *p*-values (p<=0.05), for Regulatory Area 4CDE, 2012.

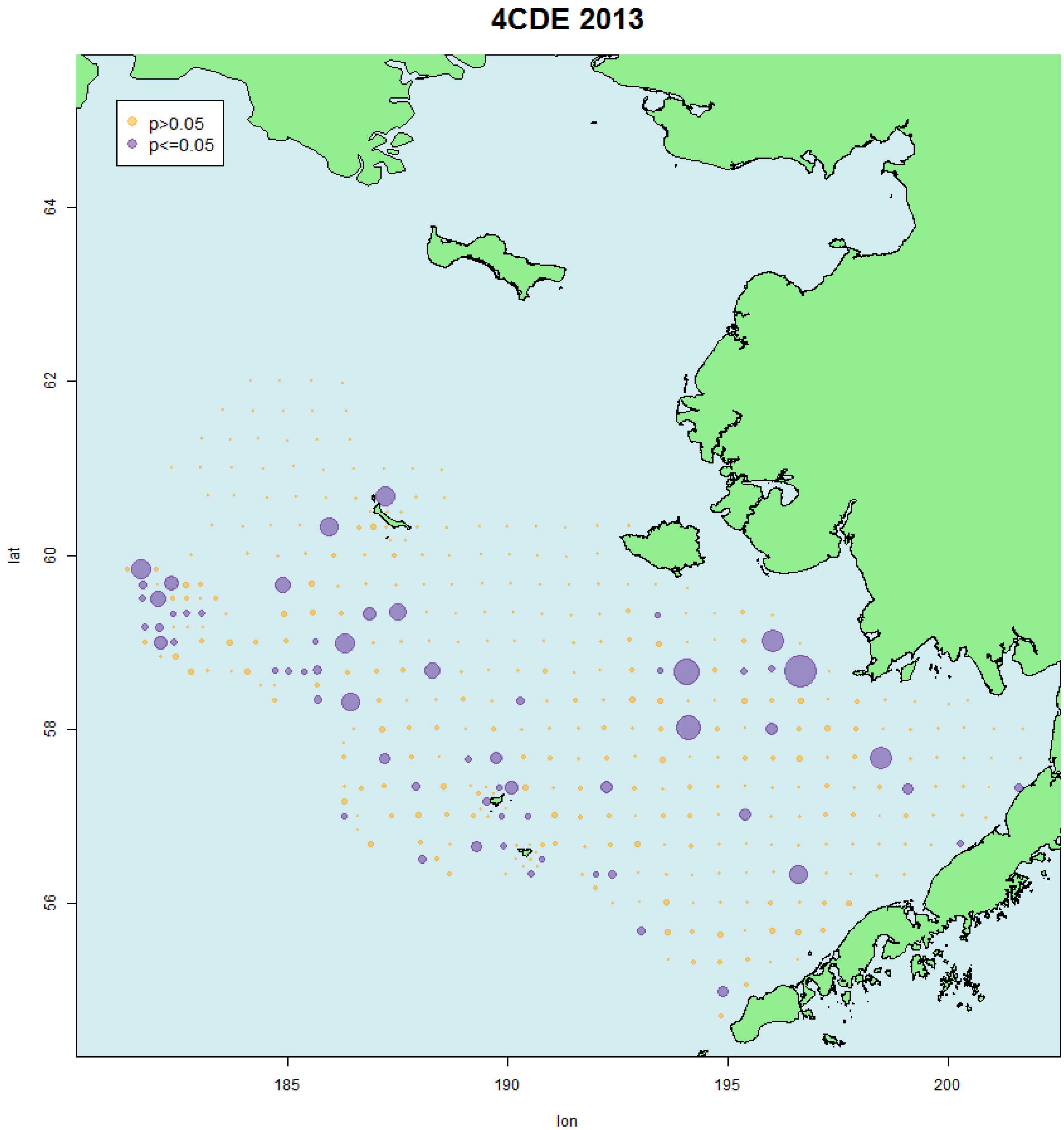


Figure 6. Maps of FISS stations with symbols showing relative sizes of discrepancy measures, *T*, and colours distinguishing those with very low *p*-values (p<=0.05), for Regulatory Area 4CDE, 2013.

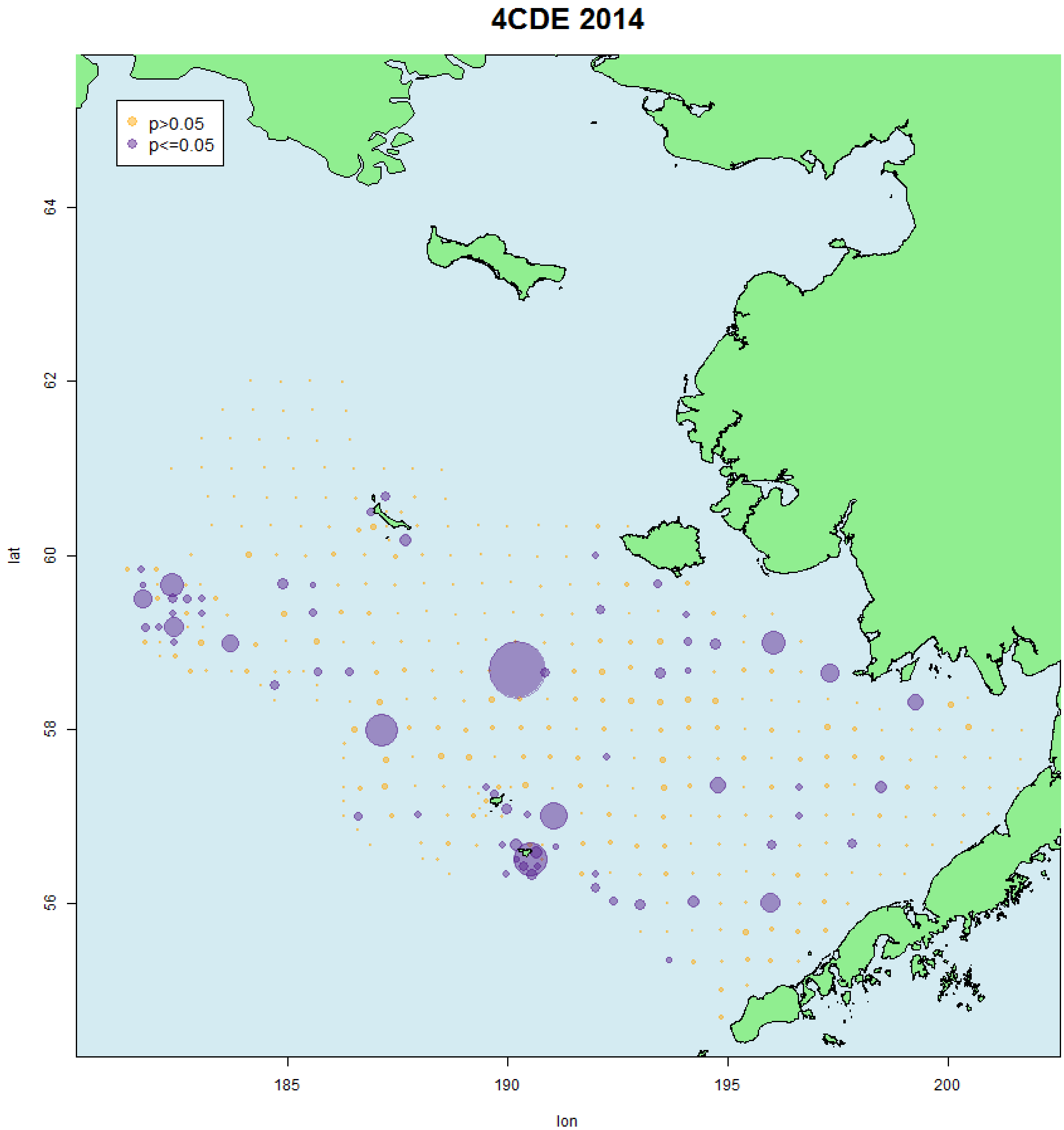


Figure 7. Maps of FISS stations with symbols showing relative sizes of discrepancy measures, *T*, and colours distinguishing those with very low *p*-values ($p<=0.05$), for Regulatory Area 4CDE, 2014.

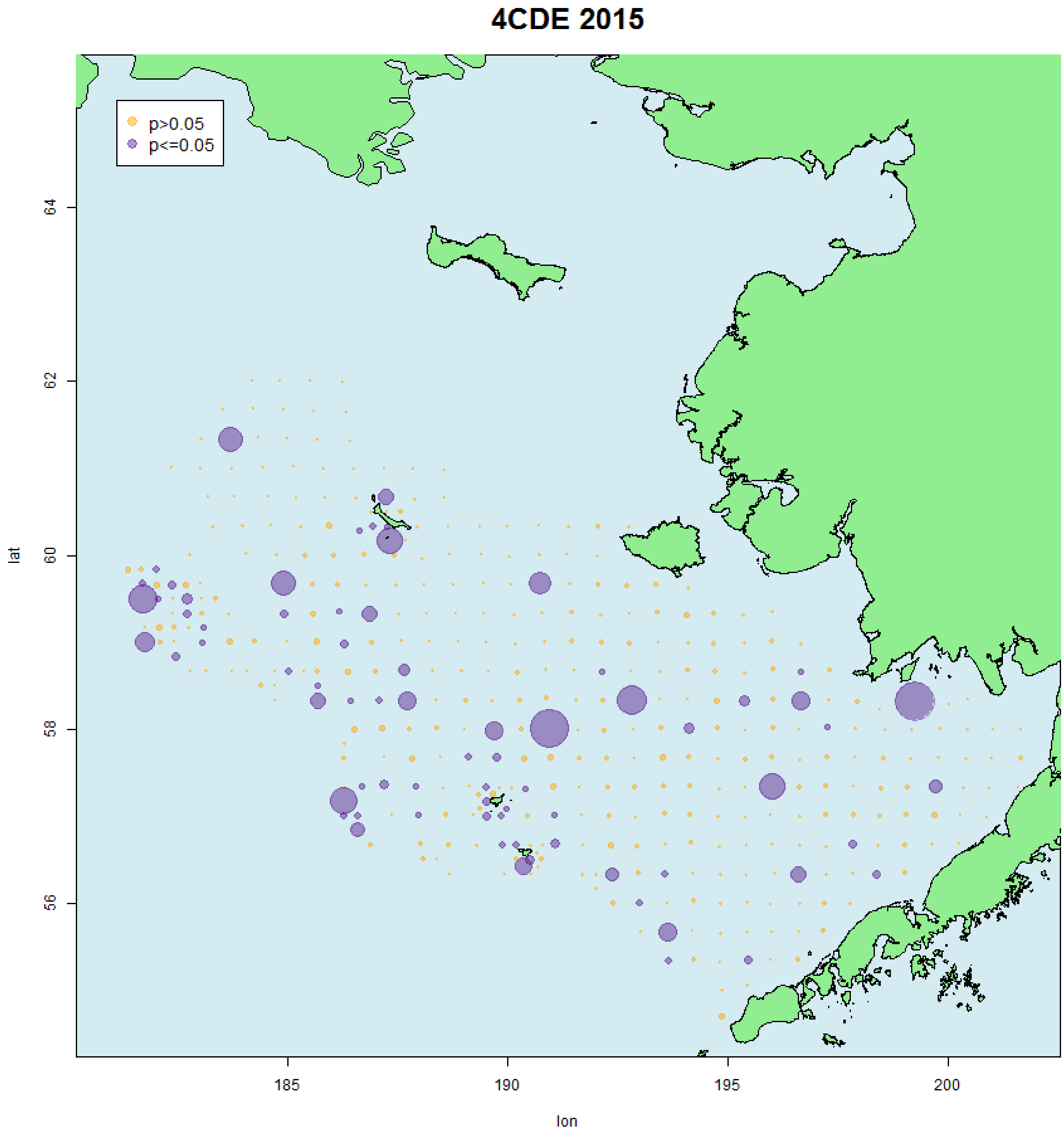


Figure 8. Maps of FISS stations with symbols showing relative sizes of discrepancy measures, *T*, and colours distinguishing those with very low *p*-values ($p<=0.05$), for Regulatory Area 4CDE, 2015.

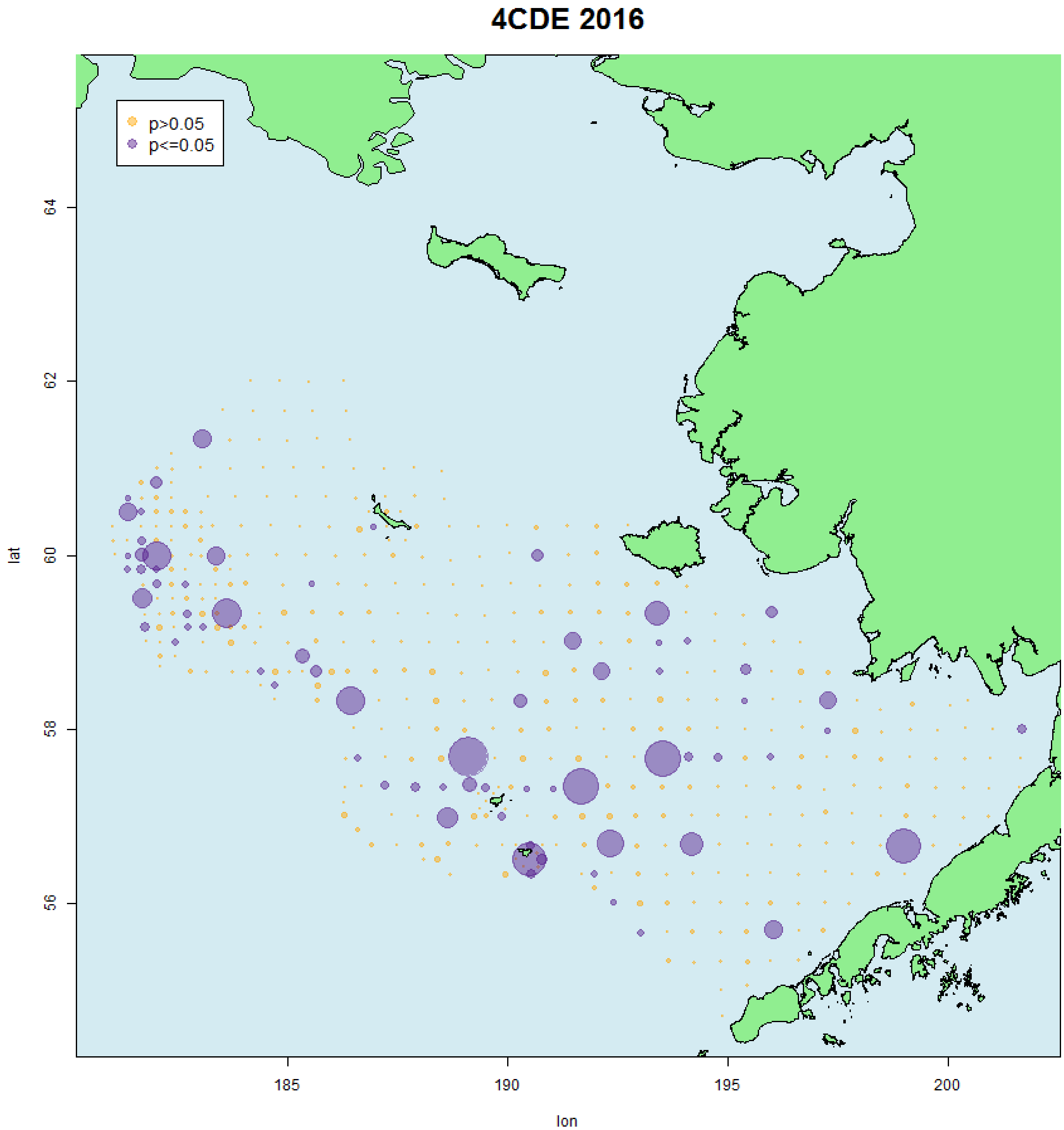


Figure 9. Maps of FISS stations with symbols showing relative sizes of discrepancy measures, *T*, and colours distinguishing those with very low *p*-values ($p<=0.05$), for Regulatory Area 4CDE, 2016.

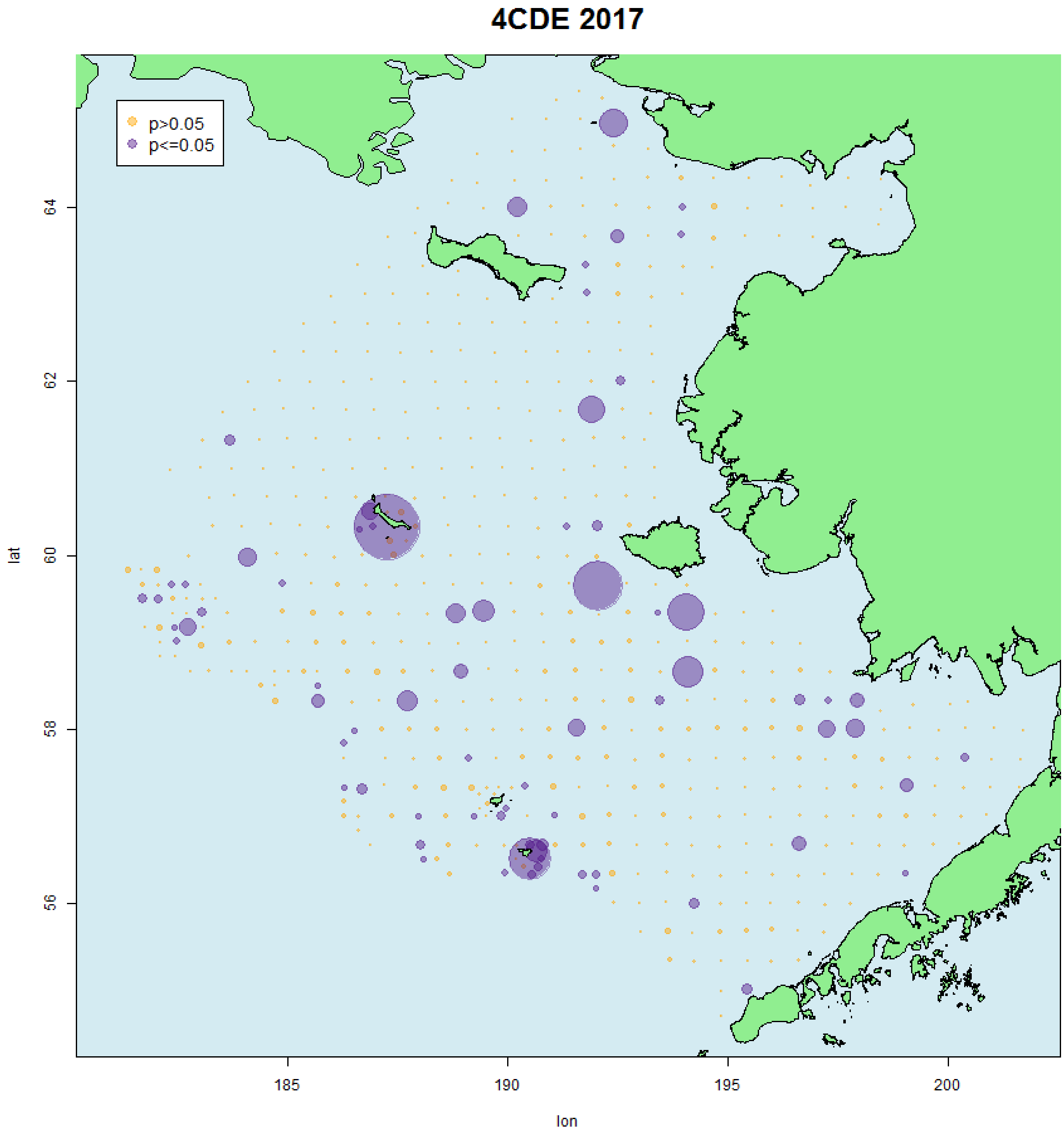


Figure 10. Maps of FISS stations with symbols showing relative sizes of discrepancy measures, *T*, and colours distinguishing those with very low *p*-values ($p<=0.05$), for Regulatory Area 4CDE, 2017.

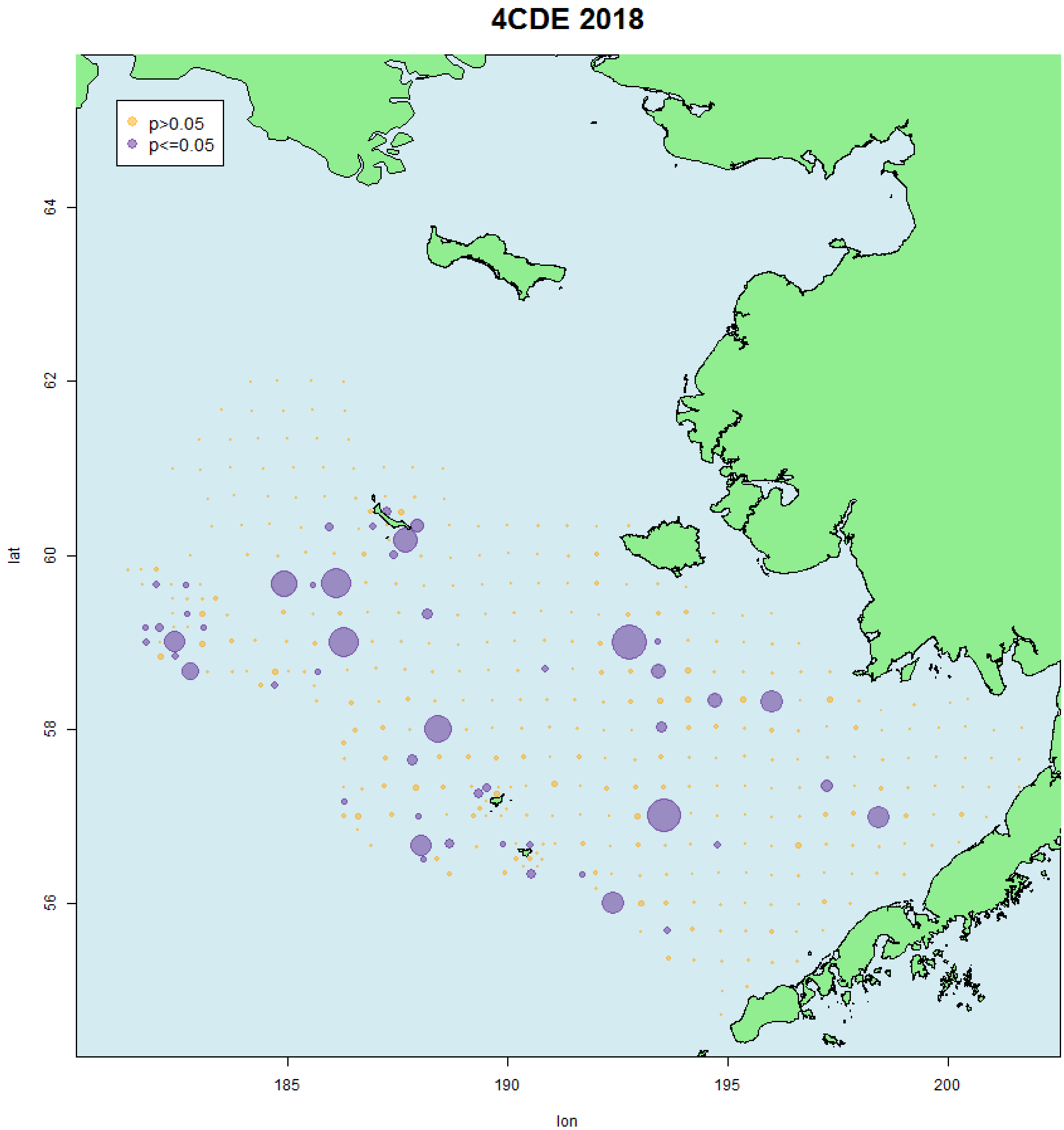


Figure 11. Maps of FISS stations with symbols showing relative sizes of discrepancy measures, *T*, and colours distinguishing those with very low *p*-values ($p<=0.05$), for Regulatory Area 4CDE, 2018.