Comisión Interamericana del Atún Tropical
Inter-American Tropical Tuna Commission




CIAT
IATTC

IDENTIFYING AND CORRECTING THE PURSE-SEINE FLEET CATCH FOR BIAS CAUSED BY THE COVID-19 PANDEMIC IN 2020-2021(SAC-13-05)

Anandamayee Majumdar, Cleridy E. Lennert-Cody, Mark N. Maunder, Alexandre Aires-da-Silva

13th Meeting of the Scientific Advisory Committee, *videoconference*, 16-20 May 2022

Outline

- Background
- Exploratory Data Analysis
- Spatio-temporal Model
- Results
- Conclusions
- Future work

Background

The problem

- The COVID-19 pandemic hindered collection of port-sampling data in 2020-2021.

- Some of the ports most affected were where bigeye tuna (BET) catch is unloaded.

- Port-sampling data are used to estimate the tropical tuna catch composition of the purse-seine fleet.

- Thus, there is concern that the Best Scientific Estimates (BSE) of catch may be biased, particularly for bigeye tuna.

Background

One possible solution

- Develop new statistical methodology that is less vulnerable to bias caused by pandemic-related data loss.

- As part of this new methodology, use other data sources for catch composition, in addition to port-sampling data.

- Require the new methodology to produce estimates as similar as possible to pre-pandemic BSEs (2010 – 2019).

- Modeling focused on developing new methodology for floating-object (OBJ) sets.

Background

A modeling challenge

- There are four primary data sources available for estimating catch composition:
  - Observer
  - Logbook
  - Cannery
  - Port-sampling

- These data sources differ in terms of:
  - Fleet coverage
  - Spatial and temporal resolution
  - Potential biases (catch amounts, species identifications)
  - Extent of data pandemic-related loss in 2020 – 2021

Exploratory Analysis

- Observer data:
  - High coverage, which was only minimally impacted by the pandemic.
  - Fine-scale space-time resolution.

- Compared port-sampling and observer species composition at their finest common resolution: 5° area x month x year.

- First considered a simple linear relationship:

$$p_{kt} = a + b\, q_{kt} + e_{kt}$$

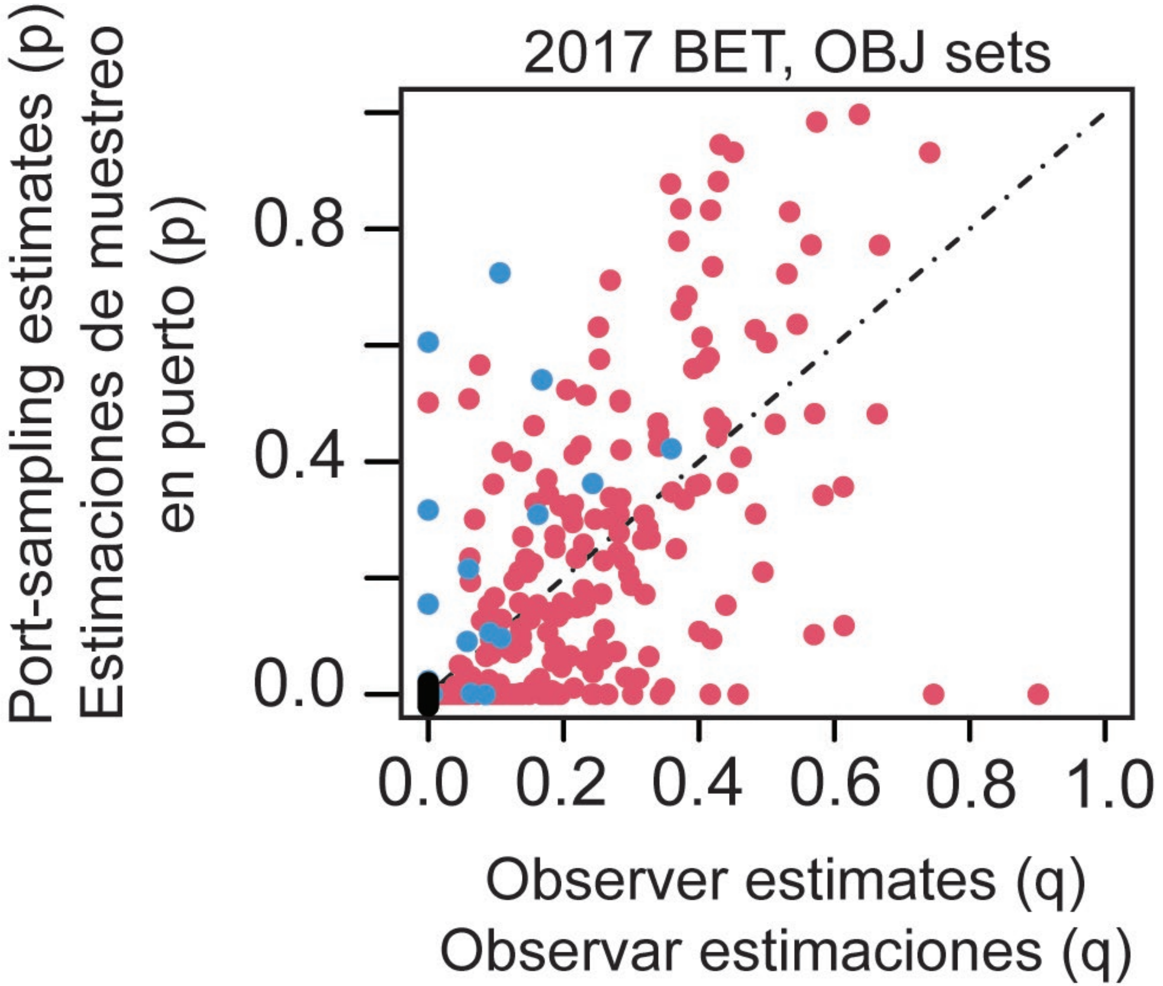
where

p_{kt} port-sampling proportions in k-th area and t-th month

q_{kt} observer data proportions in k-th area and t-th month

e_{kt} random error

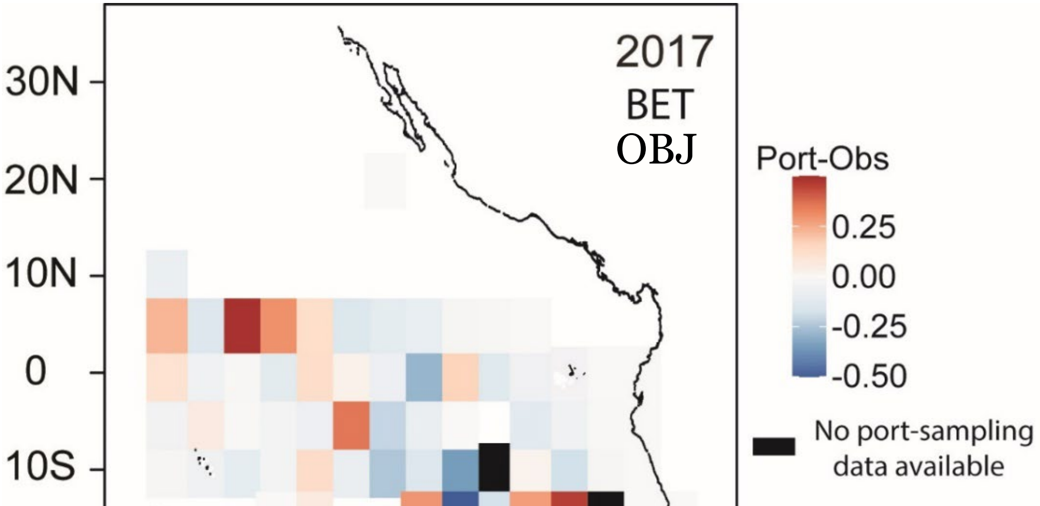
Exploratory analysis: plots of p_{kt} versus q_{kt}

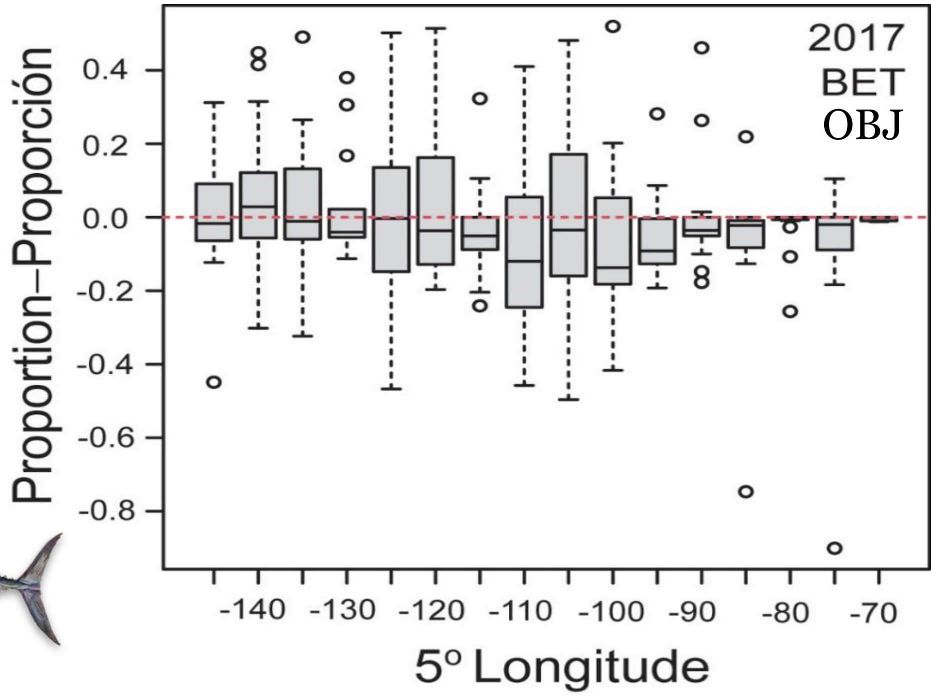


- Overall increasing relationship, generally consistent across 2010-2019.
- Variability may be due to several factors, including spatial and temporal effects.

- Black dots DEL
- Red dots OBJ
- Blue dots NOA

Exploratory analysis: spatial distribution of $(p_{kt} - q_{kt})$





- Significant spatial pattern – the medians (centers) of the 5° longitudinal box-and-whisker plot summaries change as a function of longitude.
- Significant spatial variation – the interquartile ranges and the ranges of the box-and-whisker plots change with the longitude.
- Inter-annual fluctuations.

CIAT IATTC

Exploratory analysis: spatial regression

- Multiple regression models fitted for each year:

$$log(p_{k.}) = a + b\, log(q_{k.}) + e_{k.}$$

$$log(p_{k.}) = a + b\, log(q_{k.}) + lat_{k.} + lon_{k.} + lat_{k.}\, lon_{k.} + e_{k.}$$

where

k. : proportions pooled over months of the year.

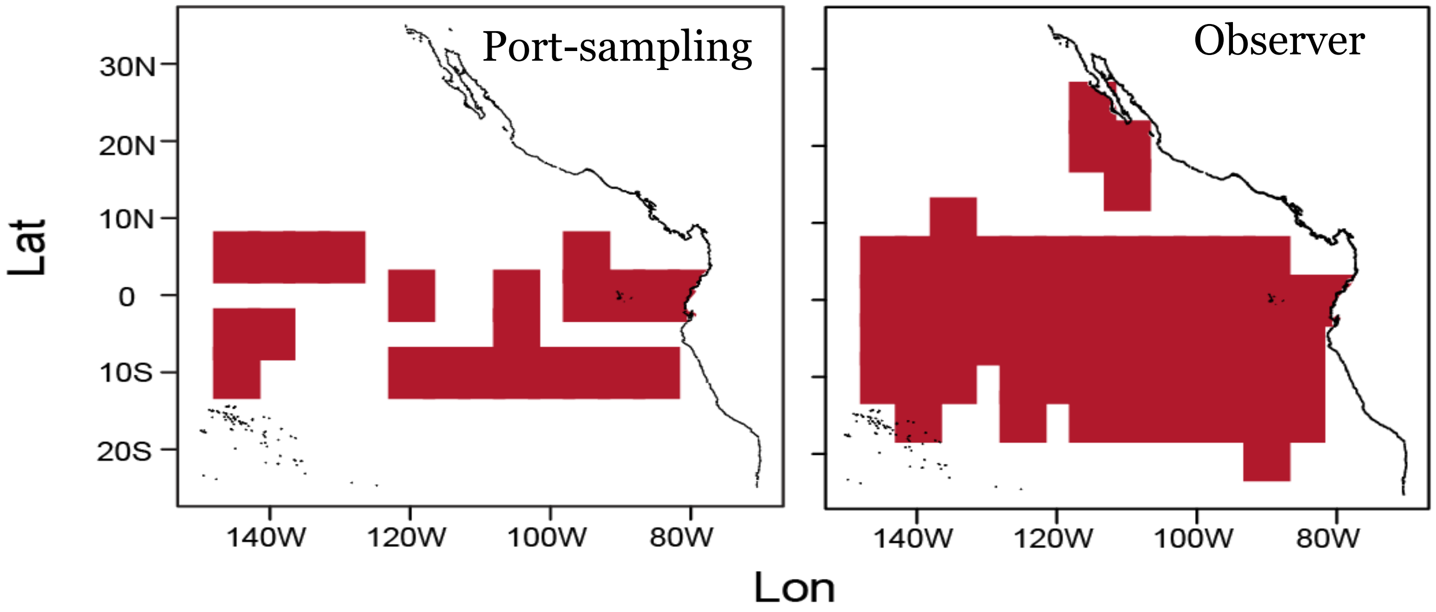
Log transformation: variance stabilizing; changes response variable range.

- All models significant.
- Models with spatial terms had better fit.
  - BET OBJ 2017
    - Non-spatial model: 29% variance explained
    - Spatial model: 45% variance explained

	Estimate	Std Error	p-value
Intercept	-7.29	0.72	< 2e-16
log(q_{kt})	0.55	0.07	3.54E-13
lat_{kt}	0.29	0.05	5.30E-06
$long_{kt}$	-0.04	0.005	1.52E-15
$lat_{kt}long_{kt}$	0.002	0.0005	3.22E-06

CIAT IATTC

Spatio-temporal Model

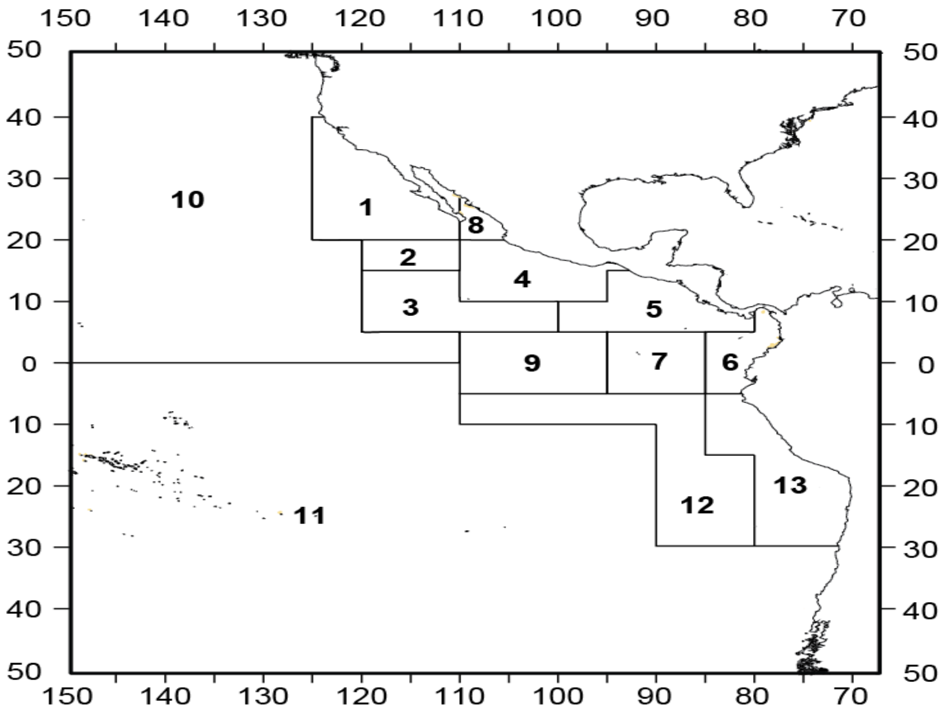


The 5° areas shown in red denote positive catch of any of the tropical tuna species

- Data sparsity is a problem.
- Data aggregated in space and time to compensate for low sample sizes.
- Spatio-temporal Conditionally Auto Regressive (CAR) model used.
- By incorporating data from multiple time periods into one model, take advantage of spatial pattern evolving in a correlated manner through time to help mitigate pandemic-related data loss.

CIAT IATTC

Spatio-temporal model: data aggregation



- Spatial aggregation reduces variance – leading to better fitting models.
- Models with various spatio-temporal resolutions considered (SAC-13-05 appendices):
  - 5° grids x month x vessel size class category
  - 13 areas x quarter x vessel size class category
  - 13 areas x year x vessel size class category

- Why these 13 areas?
  - They are the sampling areas used in the BSE methodology.
  - One of the modeling goals: produce estimates as similar as possible to the BSEs for 2010-2019.

CIAT IATTC

Spatio-temporal Model: performance measures

- Three model performance measures were used:
  - Normalized prediction sum of squared errors
  - Percent variation explained by the model
  - Correlation with the BSE for 2010-2019

- Based on these measures, the best model fit was obtained for data aggregated at the coarsest resolution (13 areas x year x vessel size class category).

Spatio-temporal model: best CAR Model for p_{kt}

- Model formulation:

$$log(p_{kt}) = Y_{kt}$$

$$Y_{kt} \mid \mu_{kt} \sim Normal\,(\mu_{kt}, \nu^2)$$

$$\mu_{kt} = X^T\beta + \phi_{kt}$$

- X is a matrix of covariates and β is the regression coefficient vector
- $X^T\beta = a + b_1\,log(q_{kt})$, the part of the spatial regression model from the exploratory analyses
- ν^2 denotes the residual error variance
- $\phi_t = (\phi_{1t}, \phi_{2t}, \ldots, \phi_{NT})$ are spatio-temporal random effects

CIAT IATTC

Spatio-temporal model: best CAR Model for p_{kt}

$$\Phi_t \mid \Phi_{t-1} \sim N(\rho_T \, \Phi_{t-1}, \;\; \tau^2 \, Q(W, \rho_S)^{-1})$$

$$\Phi_1 \sim N(0, \;\; \tau^2 \, Q(W, \rho_S)^{-1})$$

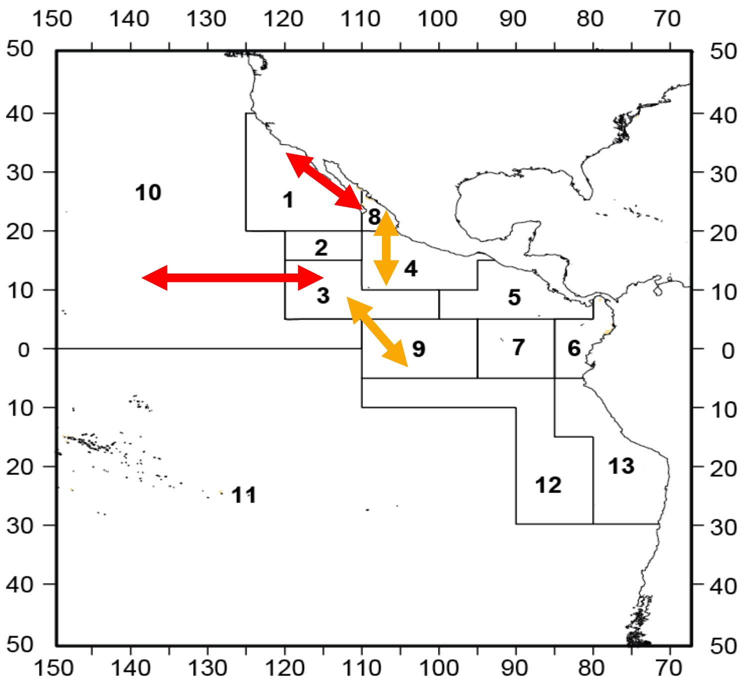
$$Q(W, \rho_S) = \;\; \rho_s \, [(diag \, W.1) - W] + (1 - \rho_S)I$$

- ϕ_{t} is an AR(1) process based on 5 consecutive years.
- temporal autocorrelation is induced via the mean, $\rho_{\mathrm{T}} \, \Phi_{t-1}$, whereas spatial autocorrelation is induced through the variance $\tau^2 \, \boldsymbol{Q}(\boldsymbol{W}, \rho_S)^{-1}$.
- Adjacency matric W= (w_{kj}) gives the weights of the strength of the spatial association between the j−th and k−th regions (W is symmetric).

CIAT IATTC

Spatio-temporal model: adjacency matrix W

- BSE methodology spatial ‘substitution’ rules: catch composition for areas without port-sampling data is based on data from ‘neighboring’ areas.
- W formulated to mimic the BSE spatial substitution rules.



First 2 rows of BSE spatial substitution matrix:

1	2	3	4	5	6	7	8	9	10	11	12	13
8	1	10	5	7	7	9	1	7	3	12	13	12
2	4	9	8	9	13	6	4	3	11	9	11	6

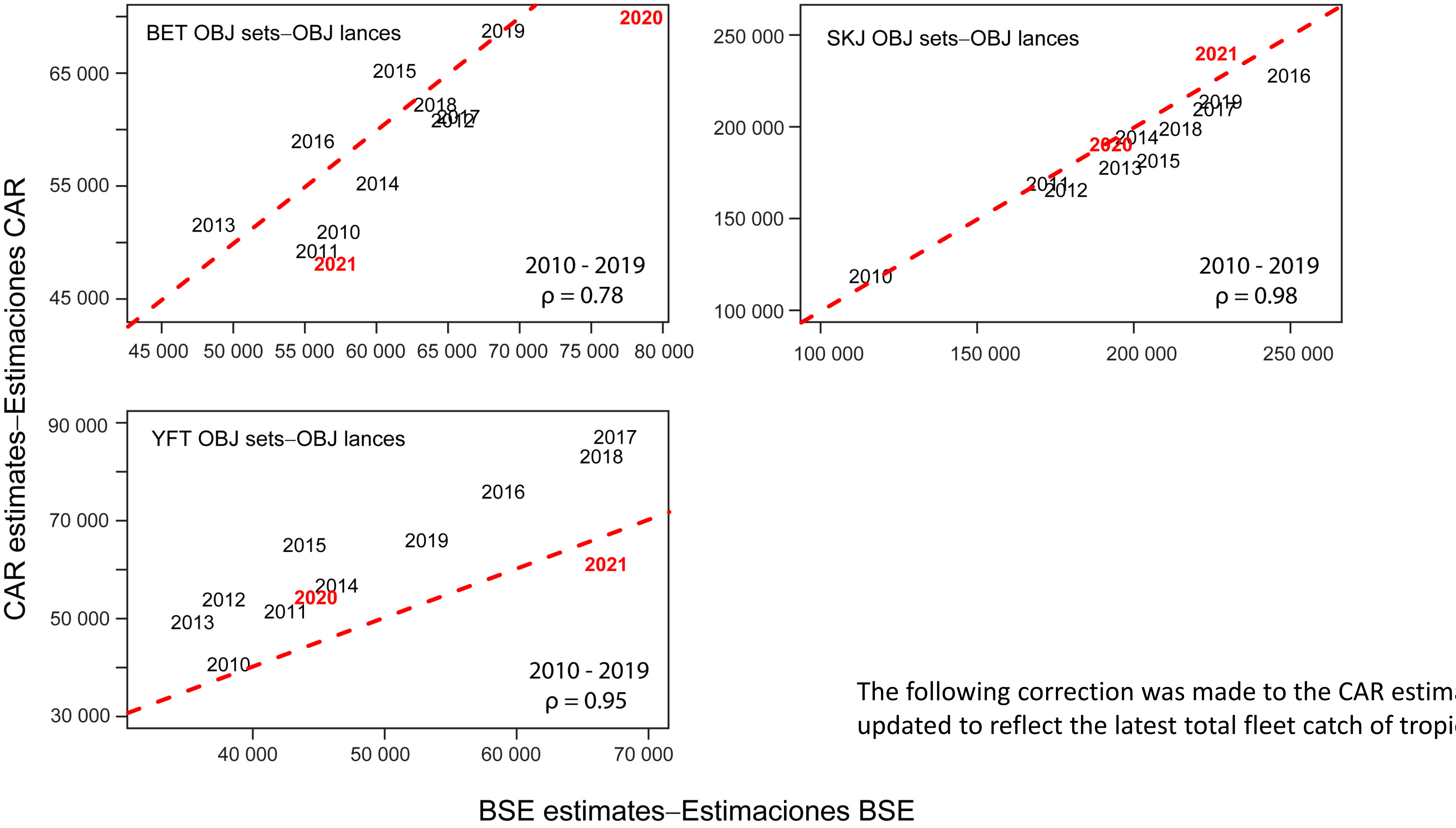
Red: 1st choice
Orange: 2nd choice

CIAT IATTC

Spatio-temporal model: catch estimation

- Estimate of total BET catch in OBJ sets
  - Species proportions estimates were obtained from the CAR model, by stratum.
  - These proportions were multiplied by the total fleet catch of tropical tunas to obtain the stratum estimates of BET.
  - Total BET is the sum of the stratum estimates.
- Similar CAR modeling and estimation was done for skipjack tuna (SKJ).
- Yellowfin tuna (YFT) catch was estimated by subtracting the sum of the SKJ and BET estimates from the total fleet catch of tropical tunas.

Results: comparison of CAR to BSE



The following correction was made to the CAR estimates shown in these figures: updated to reflect the latest total fleet catch of tropical tunas.

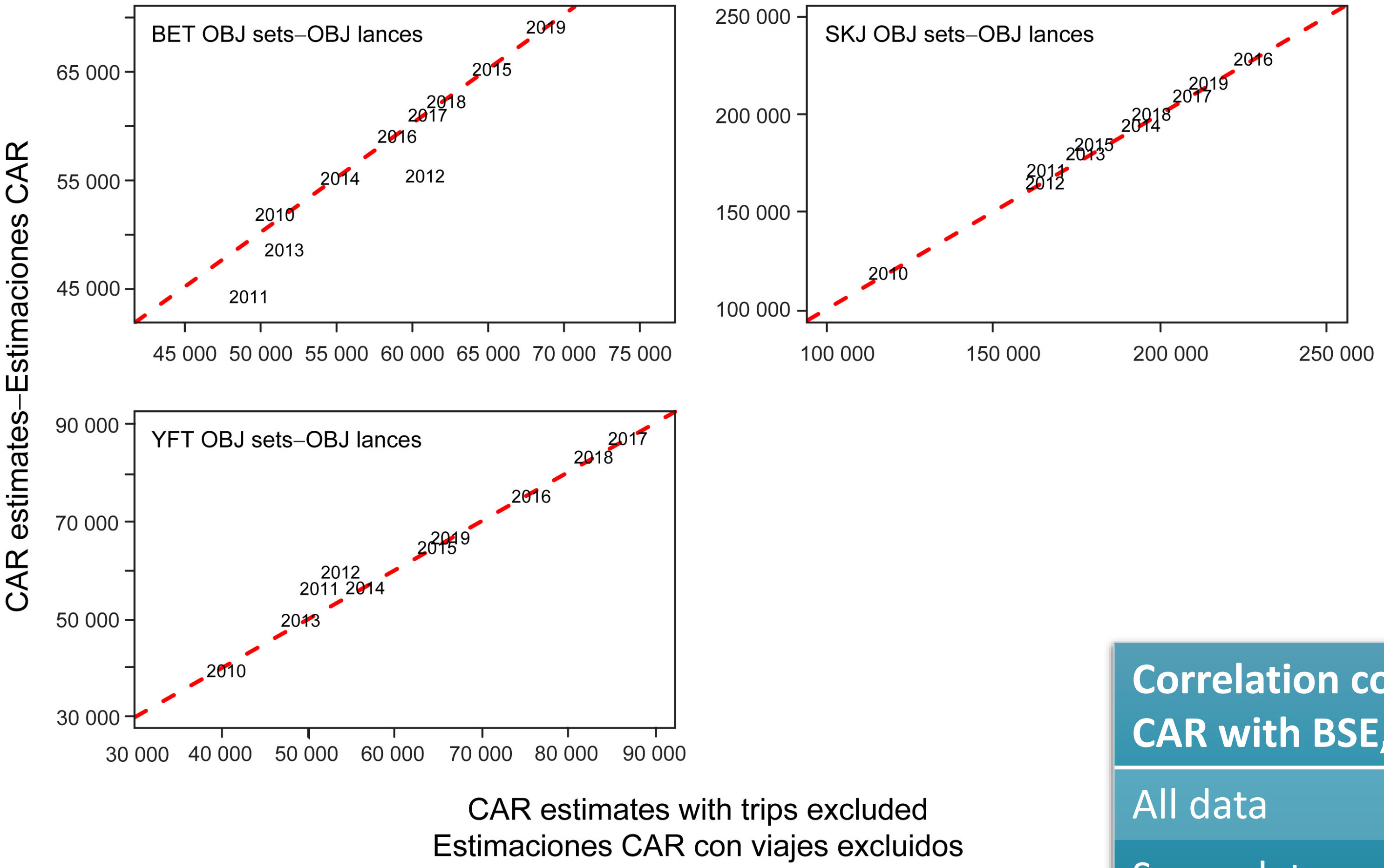
Results: estimated catch and bias for OBJ sets

Estimated values	2020 CAR*	2020 BSE	2020 Bias*	2021 CAR*	2021 BSE	2021 Bias*
BET	69 901 t	78 208 t	8 307 t (12%)	48 088 t	56 861 t	8 773 t (18%)
SKJ	190 243 t	191 399 t	1 156 t (0.6%)	239 692 t	225 132 t	- 14 560 t (- 6%)
YFT	53 924 t	44 461 t	- 9 463 t (- 18%)	60 701 t	66 488 t	5 787 t (10%)

- Bias = [BSE estimate - CAR estimate]/CAR estimate

* The following correction was made to the 2020-2021 CAR estimates: updated to reflect the latest total fleet catch of tropical tunas.

Results: sensitivity analysis for 2020



- CAR estimates were made for 2010-2019, excluding data to mimic the pandemic-related data loss in 2020 (see SAC-13 INF-L).
- These estimates were compared to the CAR estimates based on complete data.

Correlation coefficients, CAR with BSE, 2010 - 2019	BET	SKJ	YFT
All data	0.78	0.98	0.95
Some data excluded	0.73	0.98	0.92

The following correction was made to the CAR estimates: updated to reflect the latest total fleet catch of tropical tunas.

Conclusions

- The COVID-19 pandemic limited the ability of port samplers to take samples, resulting in a reduction in OBJ-set samples for 2020 and 2021 of 66%* and 35%*, respectively, compared to 2019.

- The port sampling data are used to calculate the species and size composition of the catch, and therefore play a very important role in the BSE catch estimation methodology.

- Port-sampling data collection was disrupted by the pandemic in some ports more than others and this cause bias in the BSE because some fleet segments preferentially unload in specific ports (SAC-13 INF-L).

* These percentages were corrected for arithmetic errors.

Conclusions

- Spatio-temporal (CAR) models to estimate port-sampling species proportions from observer (logbook) data with overall good performance were developed.
- Simulation results suggest the CAR model performance is robust to the type of systematic data loss that occurred in 2020.
- As compared to the CAR estimates*, the OBJ-set BSEs for BET in 2020 and 2021 represent an overestimation of about 12% in 2020 and 18% in 2021.
- The results for 2021 are preliminary, however, because the 2021 CAR estimates are based on data for 2020, which was also impacted by the pandemic.
- Further research needs to be conducted to determine the robustness of the 2021 CAR estimates.

* The following correction was made to the 2020-2021 CAR estimates: updated to reflect the latest total fleet catch of tropical tunas.

Future work

- Simulations to evaluate the robustness of the CAR model 2021 estimates will be conducted.

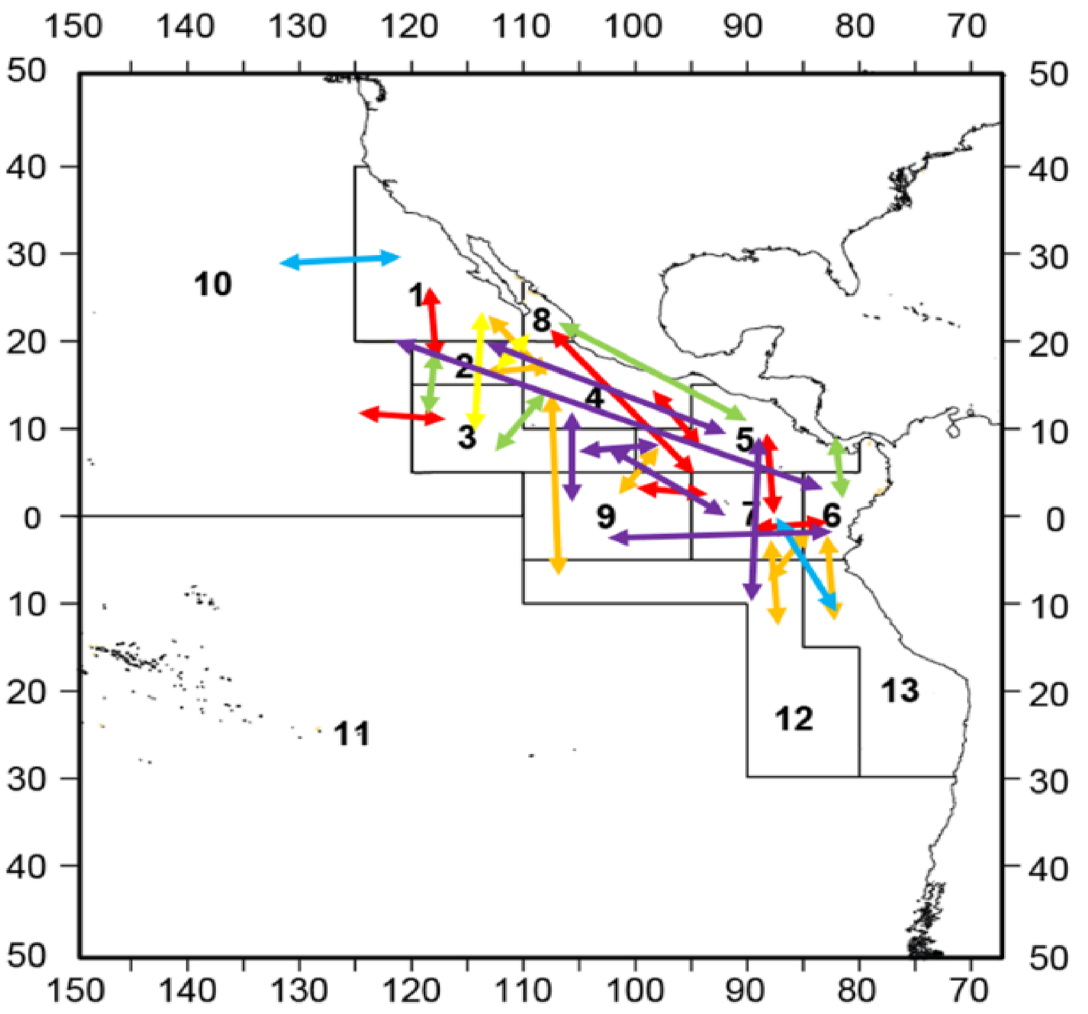
- Development of fine-scale spatio-temporal models (e.g., 5° - month or 5° - quarter) will be undertaken because the stock assessment models have a quarterly time step and the fisheries definitions differ from the 13 areas used in this analysis.

- Development of fine-scale models that are not constrained to be highly correlated with the BSE will also be undertaken.

CIAT
IATTC

Thank you! Questions?

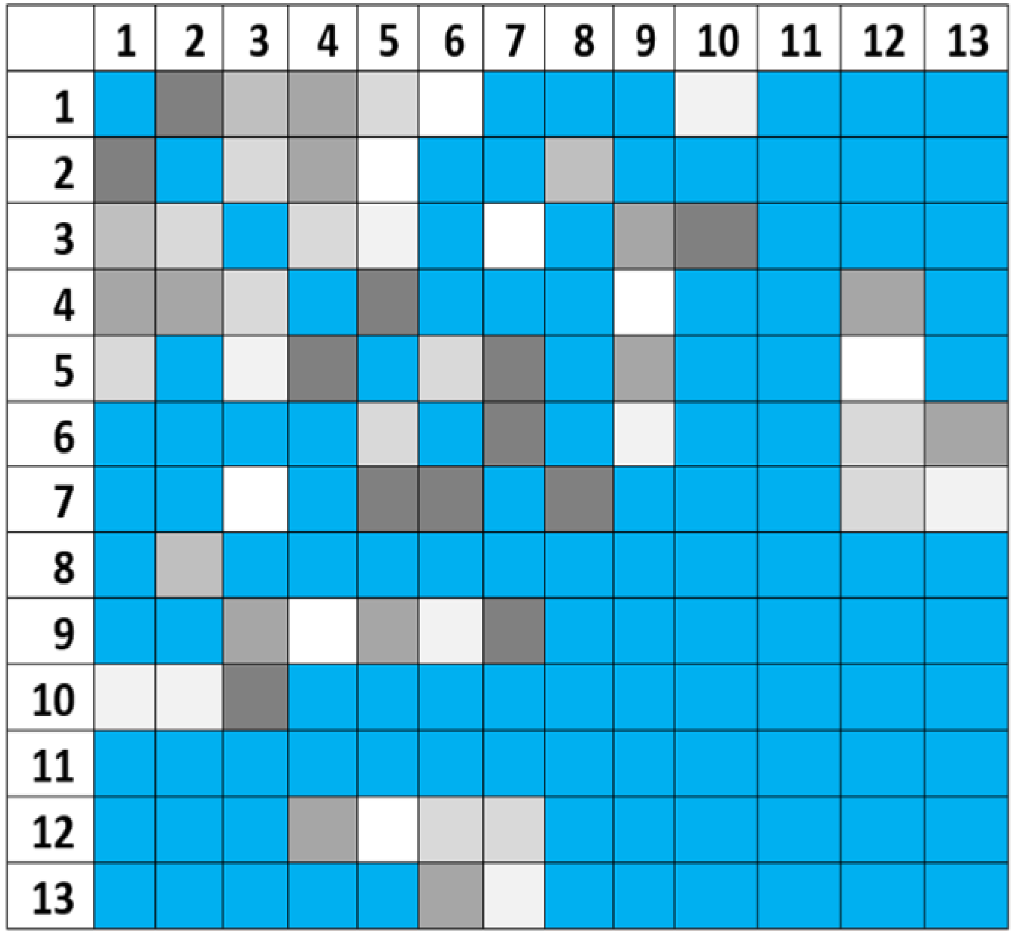
Adjacency Matrix W used in the CAR model



1	2	3	4	5	6	7	8	9	10	11	12	13
8	1	10	5	7	7	9	1	7	3	12	13	12
2	4	9	8	9	13	6	4	3	11	9	11	6

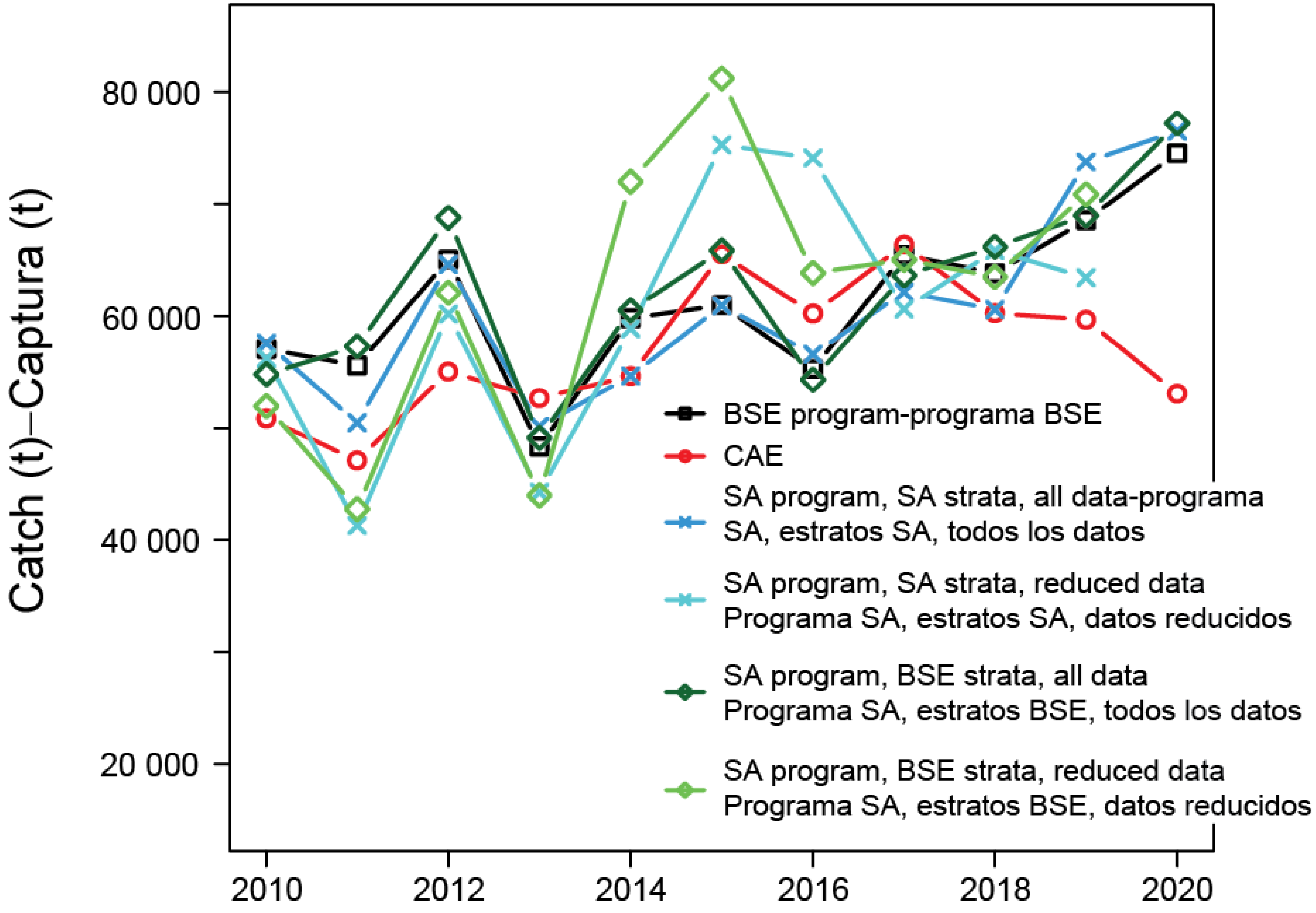
- Substitution rules are represented by colored arrows, with the rainbow color scheme.
- The substitution rules in order of the colors are *red, orange, yellow, green, blue and violet.*
- Red denotes the first order (strongest) substitution rule.
- Violet denotes the 6th order (weakest) substitution.

Adjacency Matrix W used in the CAR model



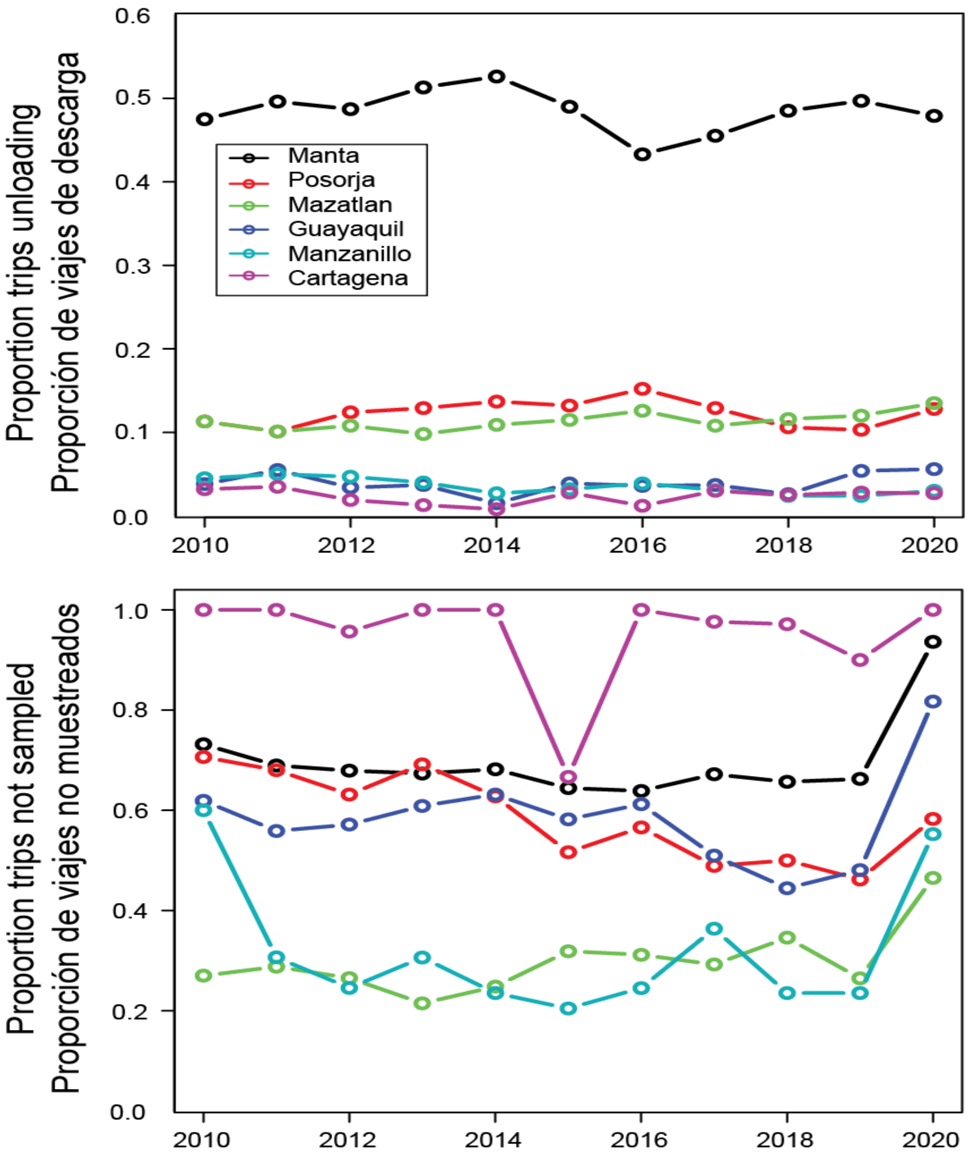
- Blue colors denote a value of zero for w_{kj} or no direct substitution.
- The black and white are shades to denote positive values between 0 and 1.
- The darker the shade, the larger the value.
- Substitution rule for BSE
  - Strata with catch but no port-sampling data based on port-sampling data from 'neighboring' strata.
  - Determined through a set of hierarchical rules (Tomlinson, 2002).

BET OBJ catch estimates for years 2010-2020



- BSE program: BSEs shown in Table A-7 of the IATTC Fishery Status Report
- SA program, BSE strata: BSEs from the stock assessment ("SA") estimation program using the BSE strata
- CAE: IATTC Catch and Effort database summary (not adjusted for coverage);
- SA program, SA strata: BSEs from the SA program using the SA strata;
- reduced data: the estimation program used the reduced port-sampling data set

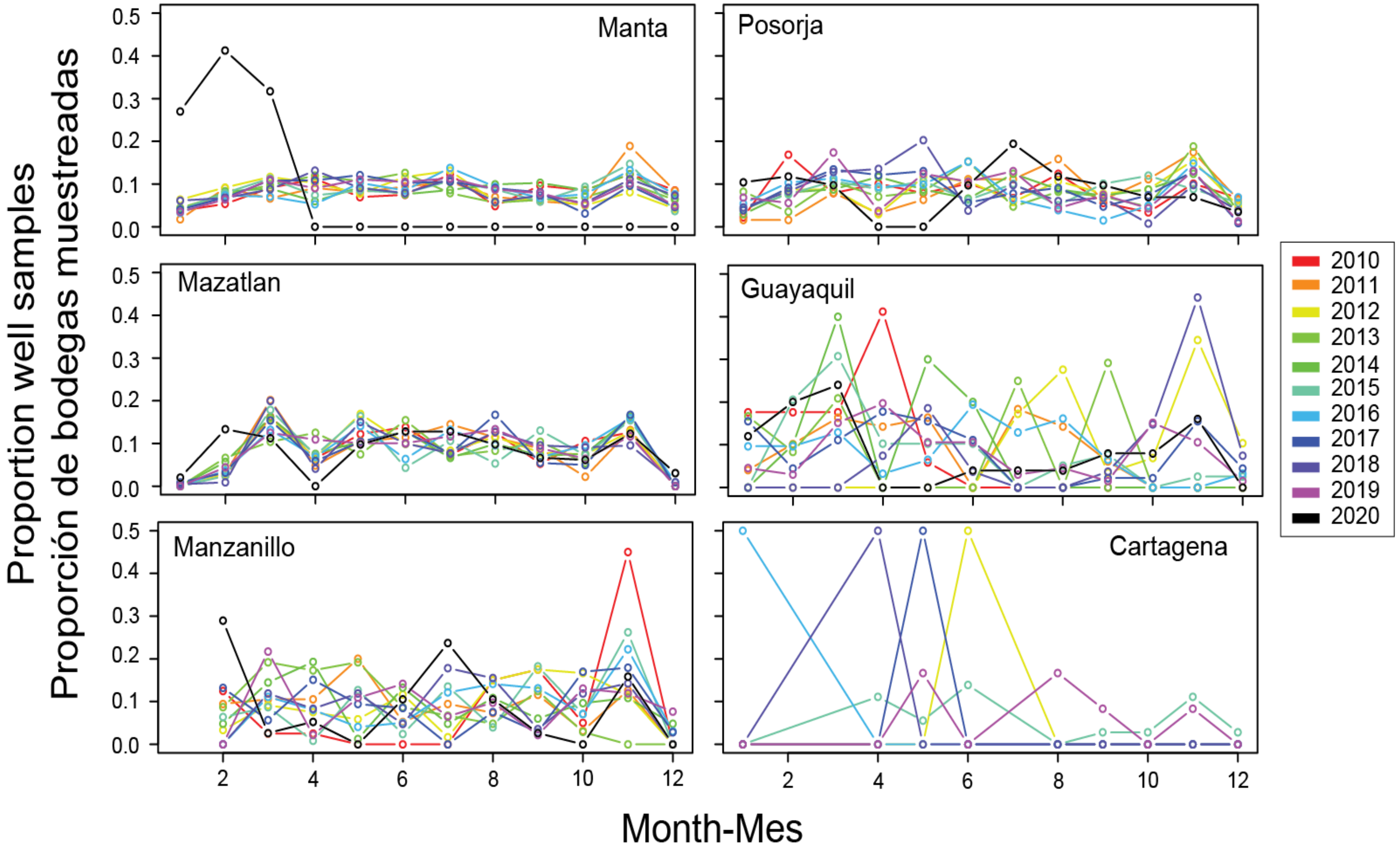
Cumulative proportions of trips that unloaded 2010-2020



- Top panel: proportion of unloadings (within a year), by port, for the 6 ports considered in this study.
- Bottom panel: annual proportion of trips *not* sampled, by port of unloading, for 2010–2020 in the 6 ports.

- Panel superior: proporción de descargas (en un año), por puerto, para los seis puertos considerados en este estudio.
- Panel inferior: proporción anual de viajes *no* muestreados, por puerto de descarga, para el periodo 2010-2020 en los seis puertos.

Proportion of wells sampled, by trip arrival month, for 2010-2020



Proportions sum to one within a year, by port.

Proporción de bodegas muestreadas, por mes de llegada del viaje, para el periodo 2010-2020 en los seis puertos. Las proporciones suman uno en un año, por puerto.

See SAC-INF-L for details