Evaluation of tag mixing assumptions in relation to spatial/temporal structure of stock assessment models: an example based on western Pacific skipjack tuna

(Custard and Doughnuts)

Dale Kolody
Simon Hoyle

SPC, New Caledonia

Outline

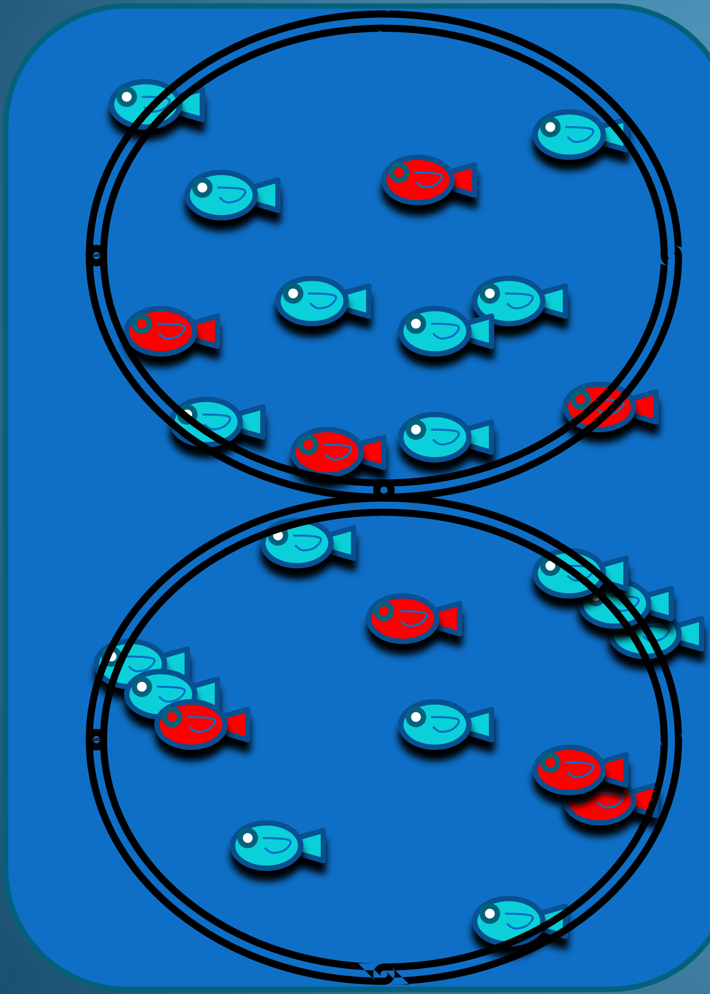
- Why is tag mixing important?
- Two analyses for examining mixing
- Application to WCPO SKJ assessment spatial structure
- A path toward formulating spatial models

A key assumption in Tag analyses:

- ***Tagged Fish are representative of the broader untagged population*** ***(or some useful subset of it)***

- e.g. Similar
  - Growth
  - M
  - Movement
  - Vulnerability to the fishery

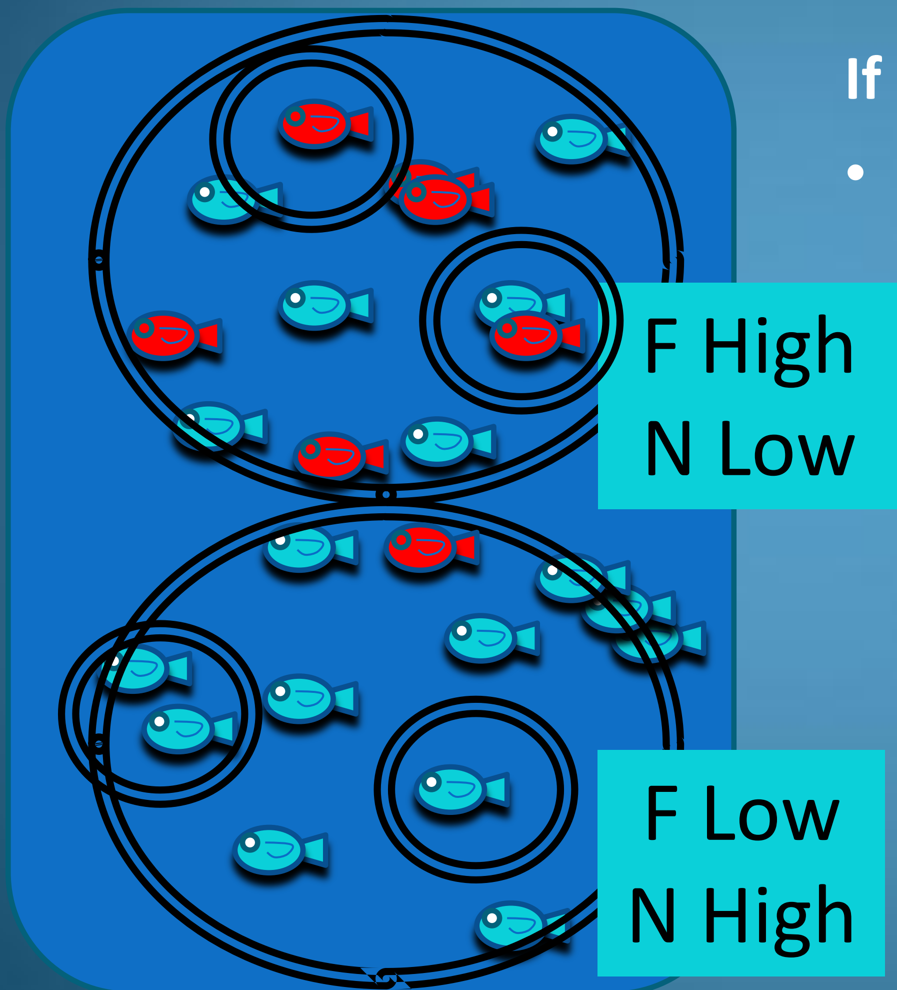
Are tags randomly mixed throughout the general population?



If mixed:

- Does not really matter where the fishery occurs

Are tags randomly mixed throughout the general population?



If not mixed:

- Tag-based estimators may be badly biased

(Unless fishery is random)

1) The CuSTaRD Analysis:

Comparison of Synchronous Tag Recovery Distributions from different release events

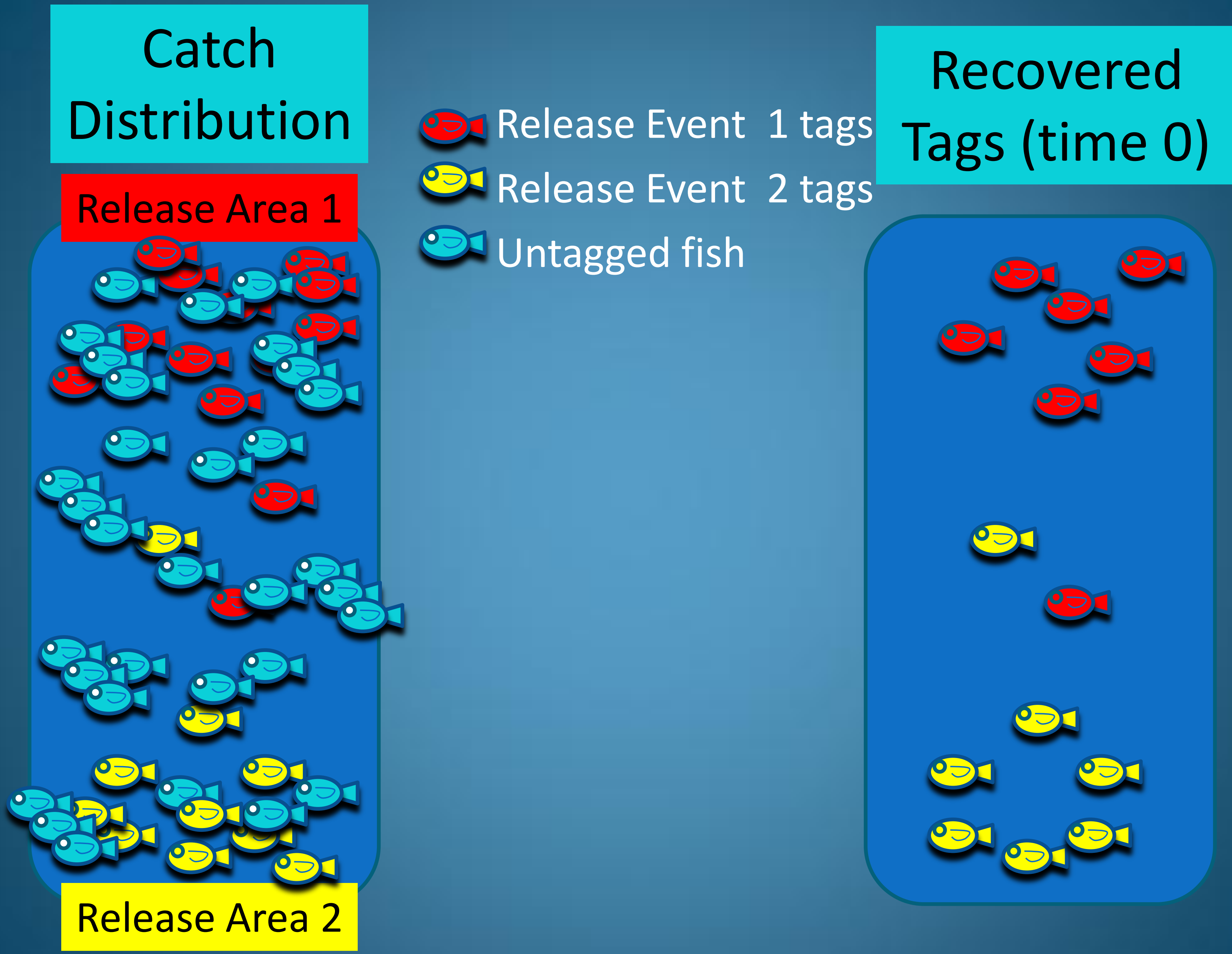
Simple Method for Examining Spatial Mixing Assumptions

- In a specific time window, compare the tag recovery spatial distributions from different release events (of similar-sized fish)

- If the distributions are significantly different, tags from the release events are not fully mixed with each other, and at least one release event is not fully mixed with the general population

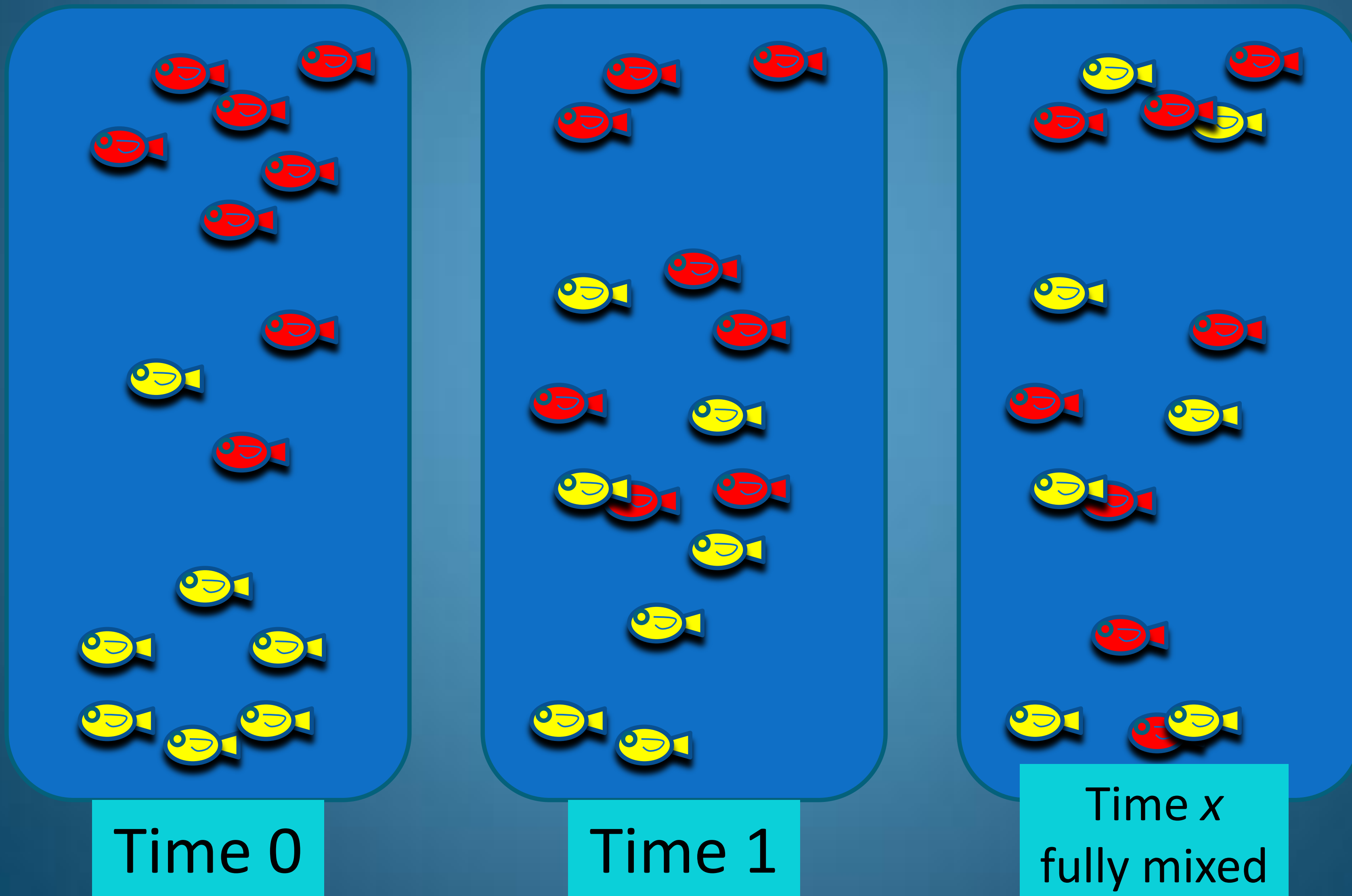
- Only interested in the shape of the tag distributions, not the number (i.e. Not worried about differing F histories)
  - Though N relevant to statistical power

- Can compare recovery distributions from release events that are separated in space or time (or both)


Catch Distribution
Release Area 1
Release Area 2
Release Event 1 tags
Release Event 2 tags
Untagged fish
Recovered Tags (time 0)

We generally assume a gradual mixing process

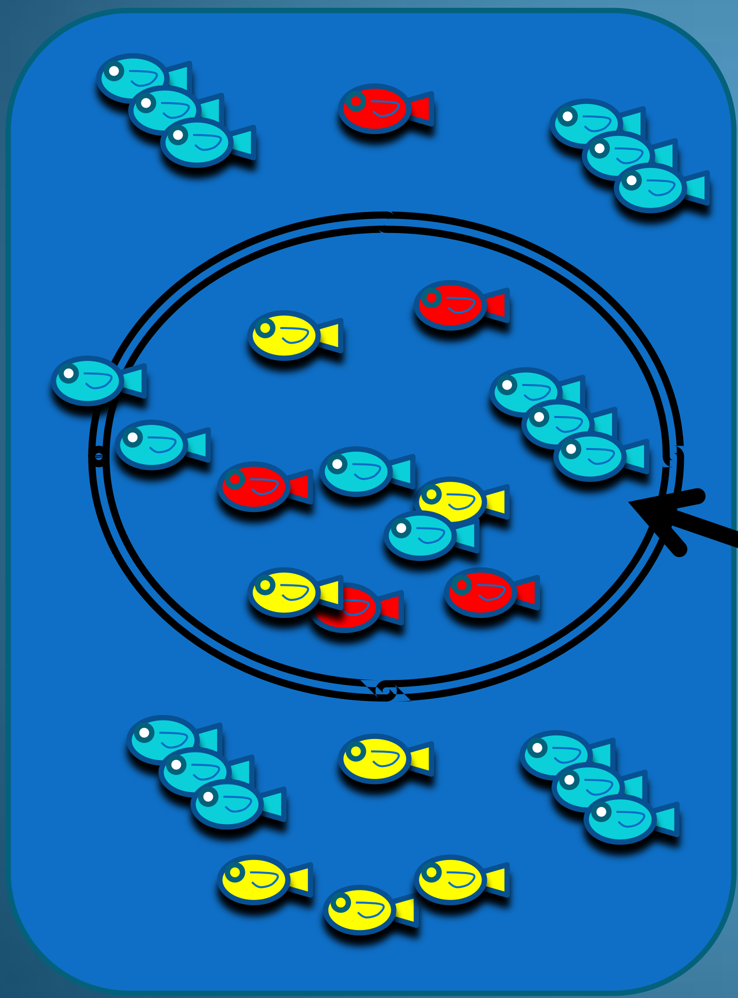
- Release Event 1
- Release Event 2



What about the catch/effort distribution and reporting rates?

- Identical for all release events in the same recovery time window
- Potentially Limits the power of the analysis.
- However, will not cause false positives (i.e. significant differences between distributions are valid regardless)

Recovery distributions might appear to be mixed when the tags are not mixed



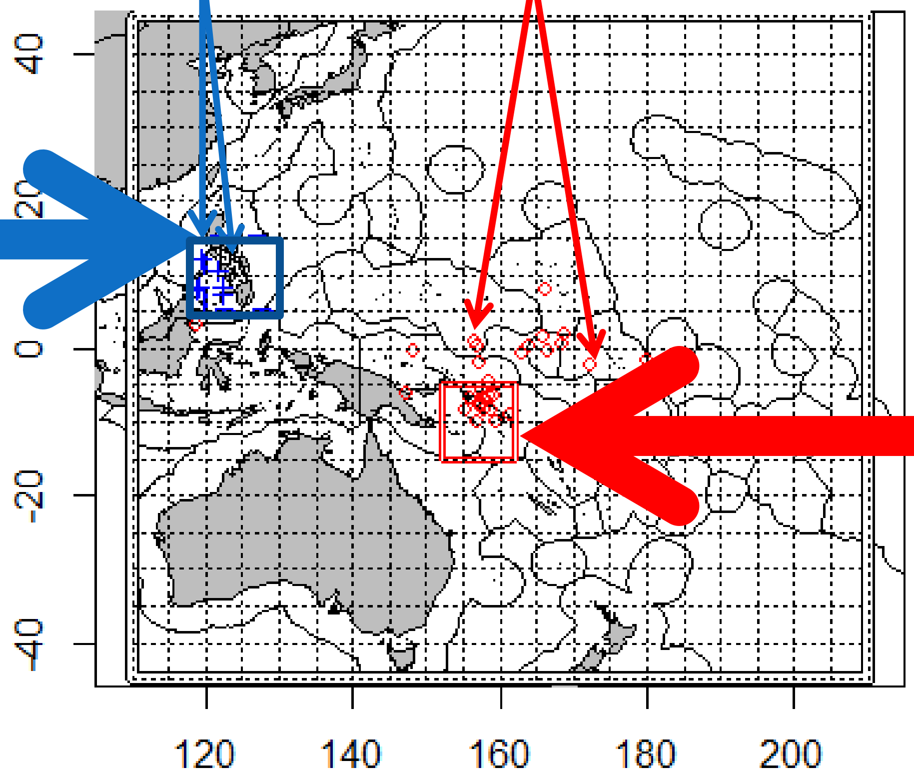
- Release Event 1
- Release Event 2
- Untagged

1) Fishery operates in a restricted area

CuSTaRD example 1

Recoveries in the same quarter

Release Event 2: Philipines 6 Qtrs previously

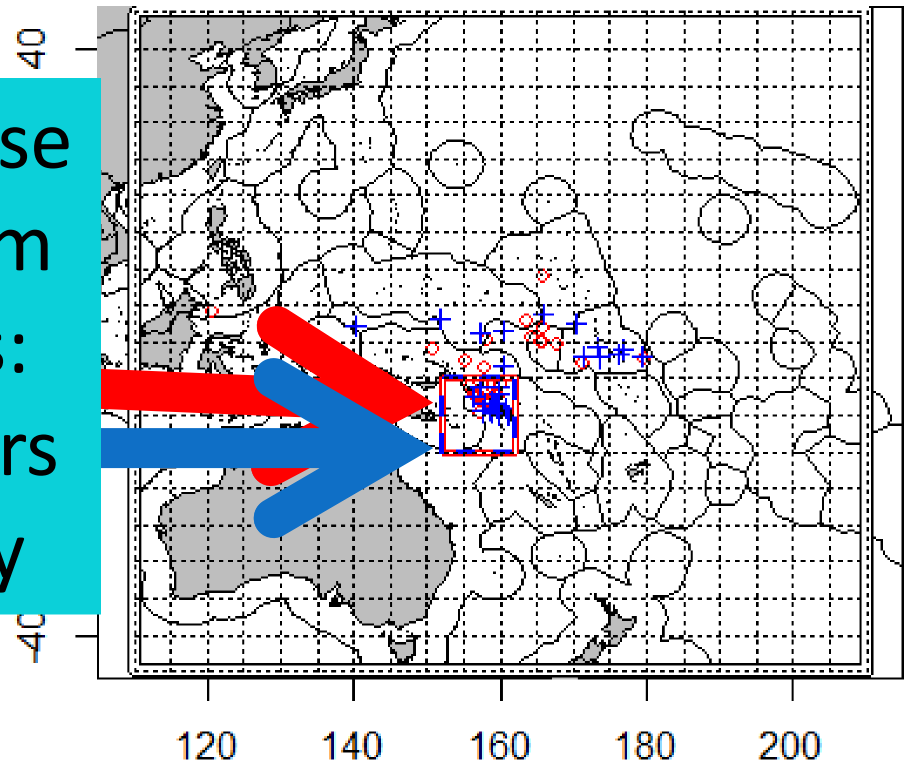


Release Event 1: Solomons 5 Qtrs previously

Recovery distributions are clearly not well-mixed

CuSTaRD example 2

Both Release Events from Solomons: 4 and 6 Qtrs previously



Recovery distributions appear similar (probably well-mixed)

How to compare 2 spatial distributions?

- P-Value as an Index of Similarity:

  1. Chi-square Contingency Table on 2D grid

  2. Nearest-Neighbour Permutation test

- Simulations indicate they are similar, but NNP more powerful with few tags

Interpreting P<0.05:

- Unlikely (5% chance) that tag distributions are the same = compelling evidence that the tags from the two release events are not well-mixed.

Can anyone think of a plausible mechanism that would cause a bias toward falsely estimating a difference in distributions?

Interpreting P>0.05:

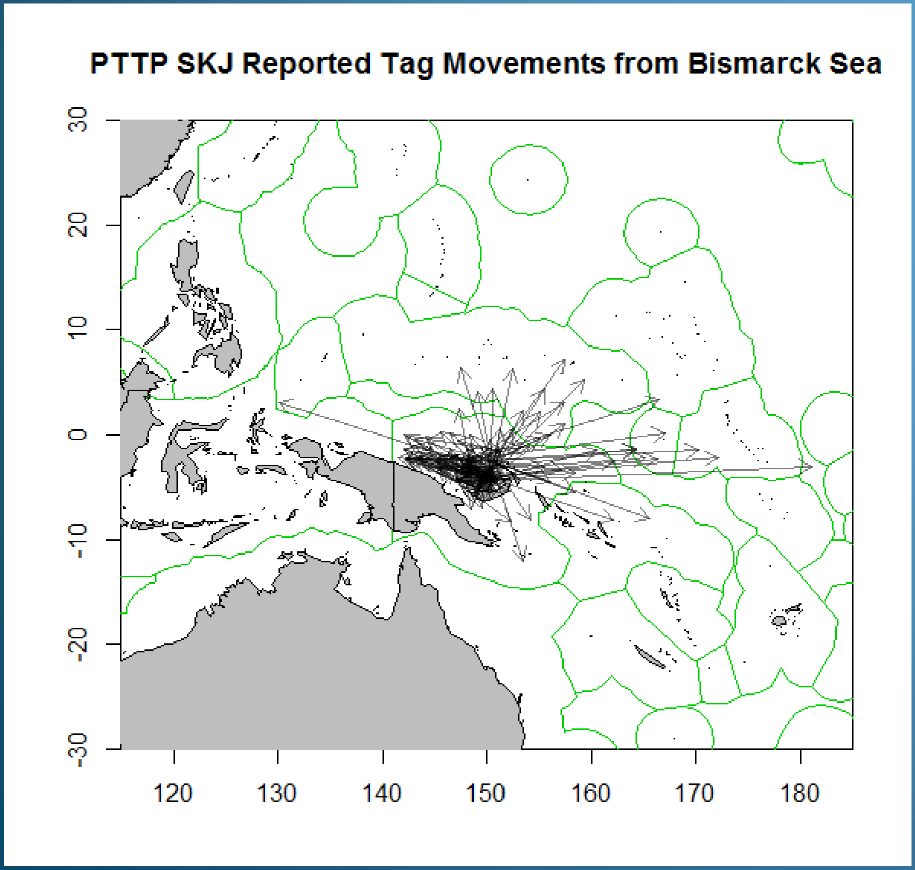
- Two release events may be well-mixed with each other and the untagged population

- Or: the release events may be well-mixed with each other but not the untagged population
- Or: Low power (few tag recoveries)
- Or: Spatially-restricted effort/tag reporting
- Or: substantial tag recovery date and position errors
- i.e. Cannot conclude that mixing is complete

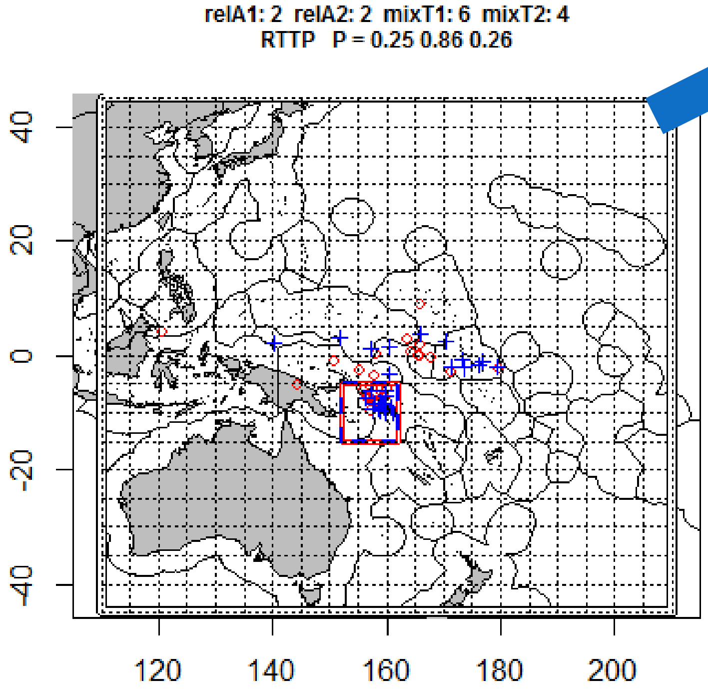
Absence of evidence that mixing is incomplete

Quick Digression: Tag Movement or Date/Position Error?

200 reported movements



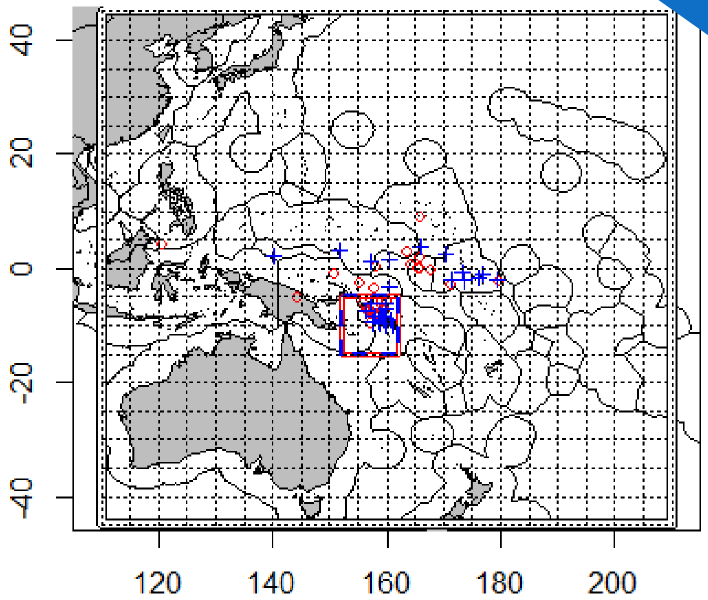
How to summarize hundreds of CuSTaRDs



Reduced to a single Index of Similarity (P-value)

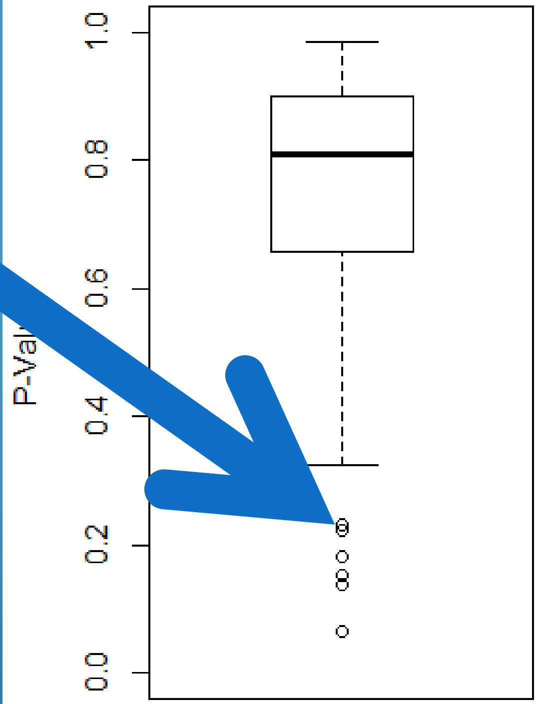
How to summarize hundreds of CuSTaRDs

P = 0.25

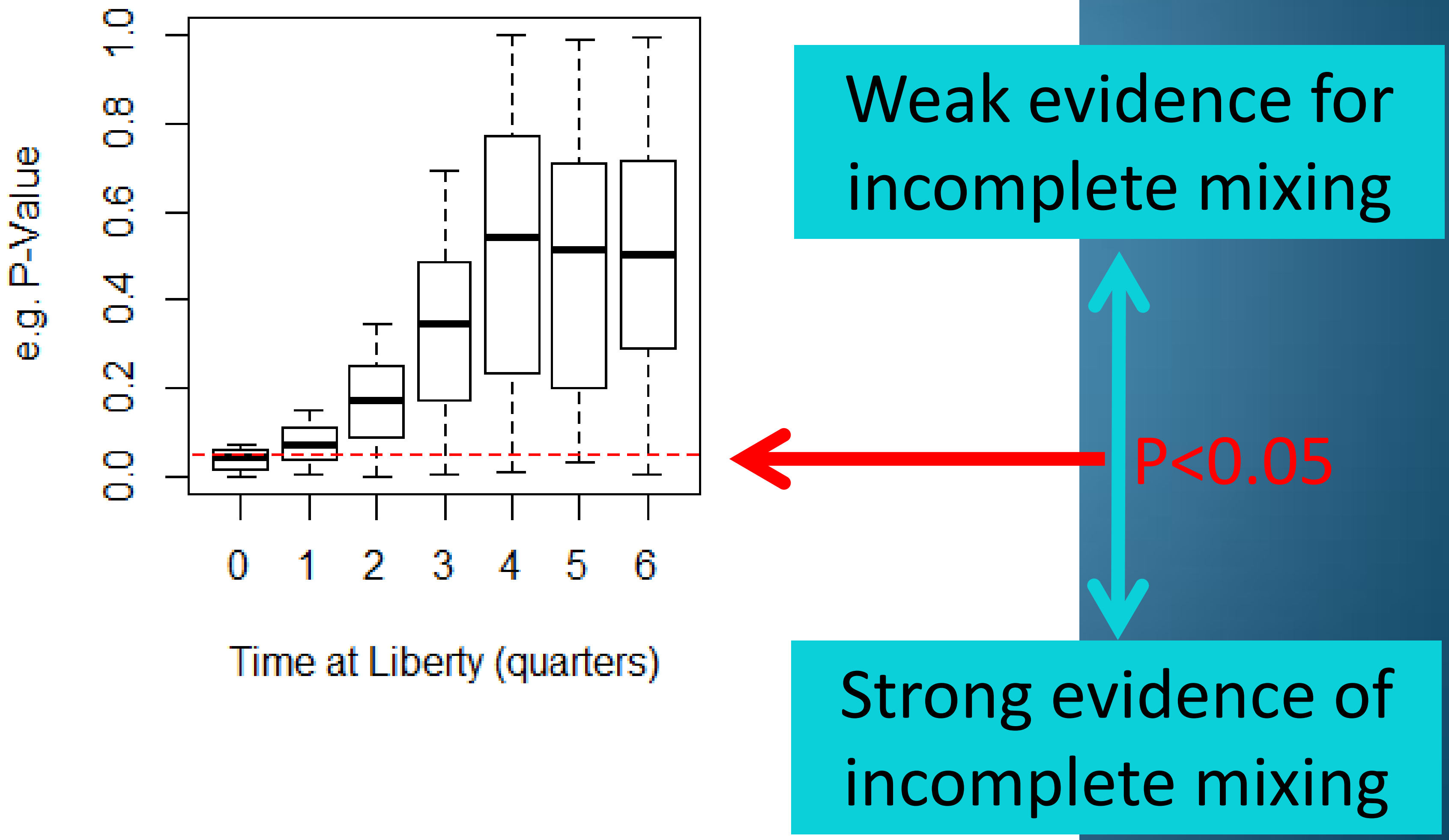


(poor evidence of non-mixing)

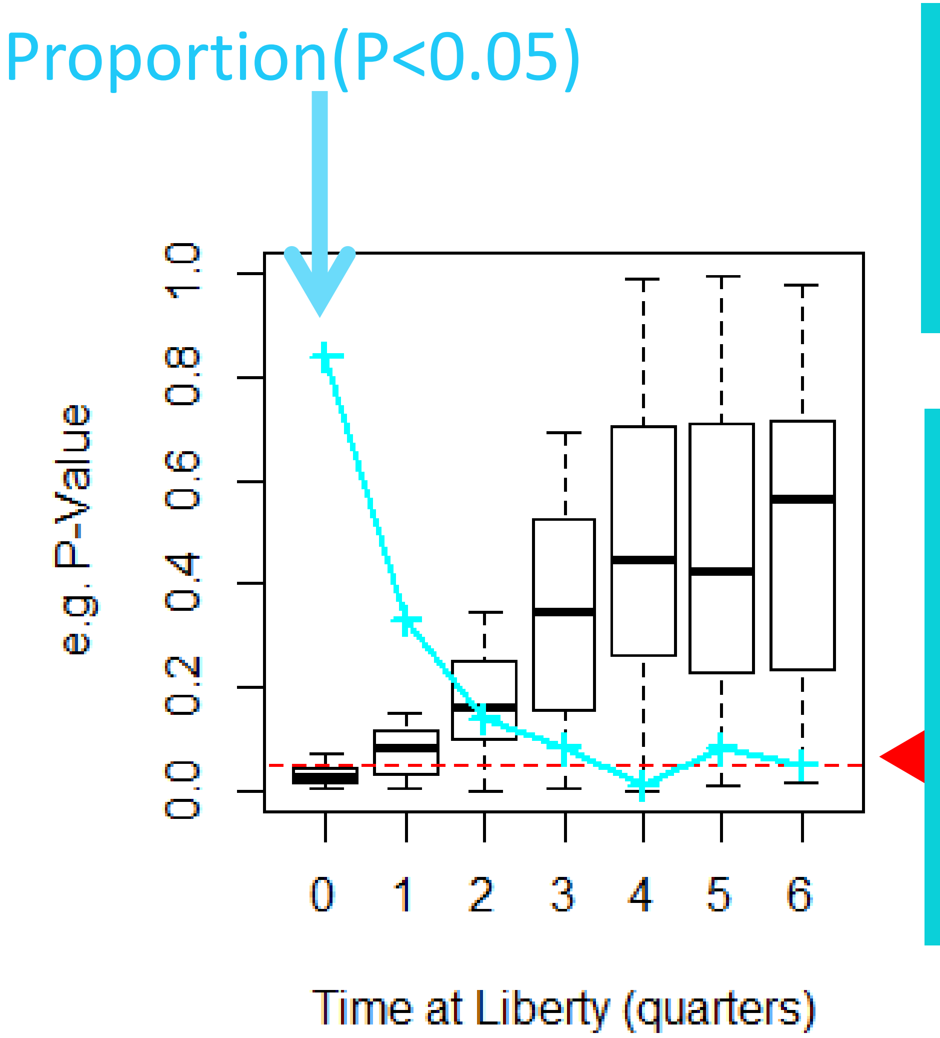
Distribution of all P-Values



Ideal Results: Evidence of non-mixing decreases over time



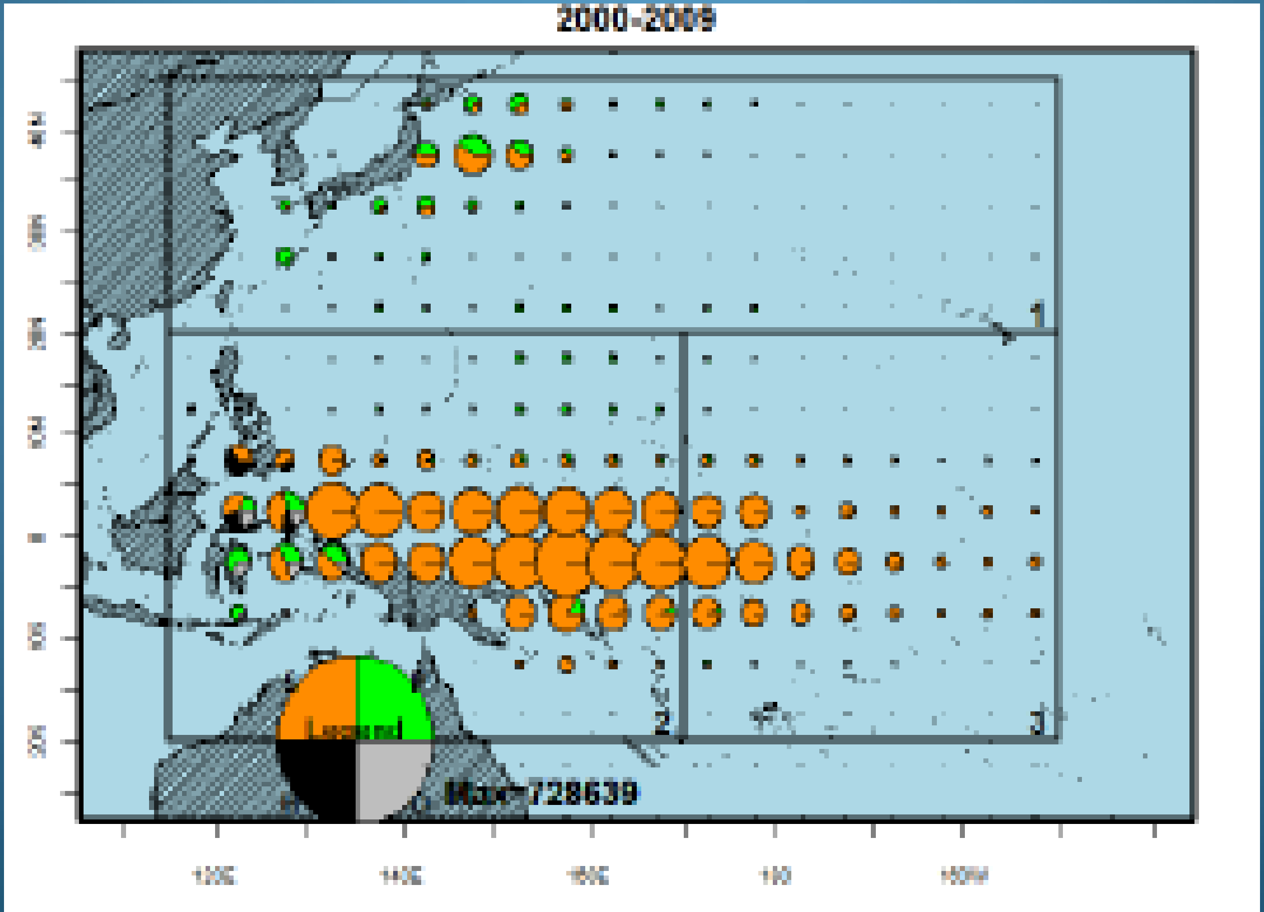
Ideal Results:

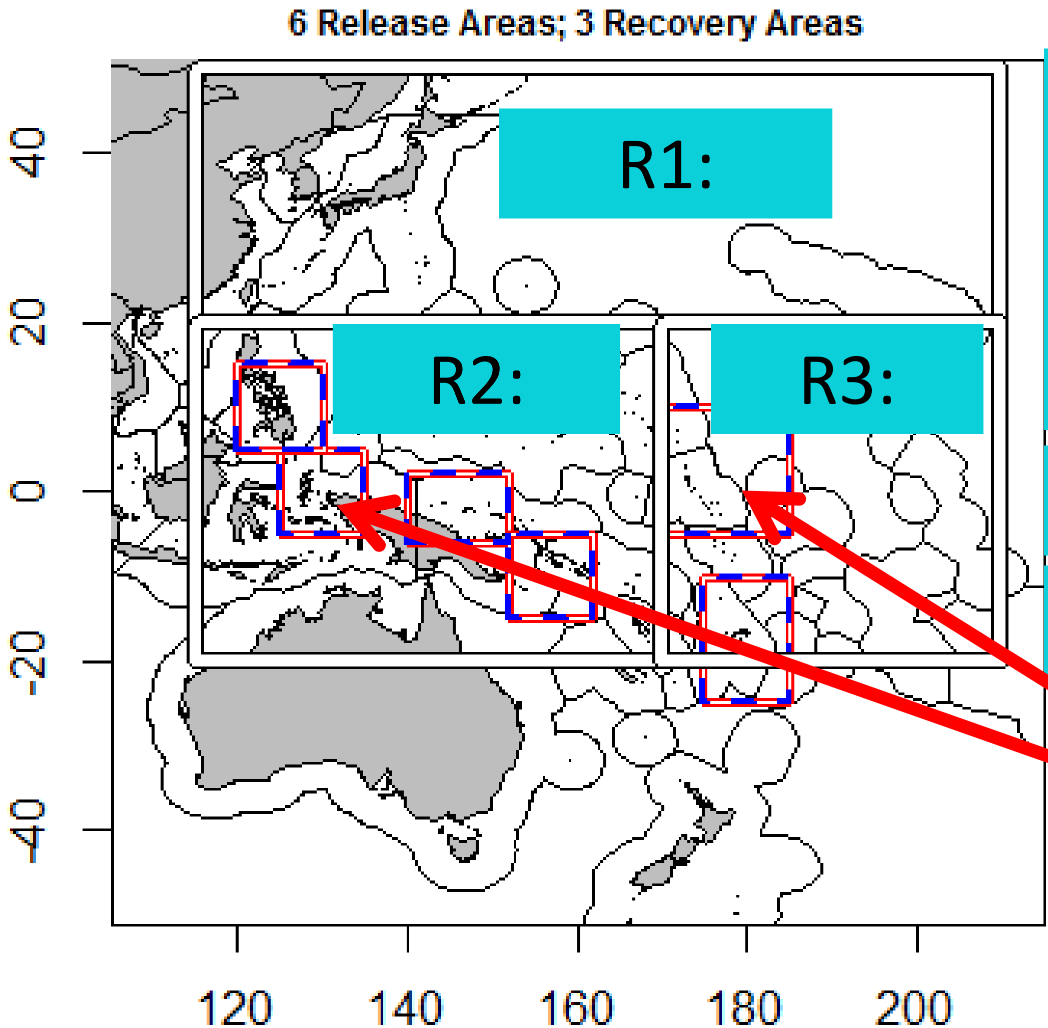


Evidence of non-mixing is strong in quarters 0-1

Reasonable to conclude that full mixing requires at least 3-4 quarters to occur

So, are tags adequately mixed in the current SKJ WCPO Stock Assessment?





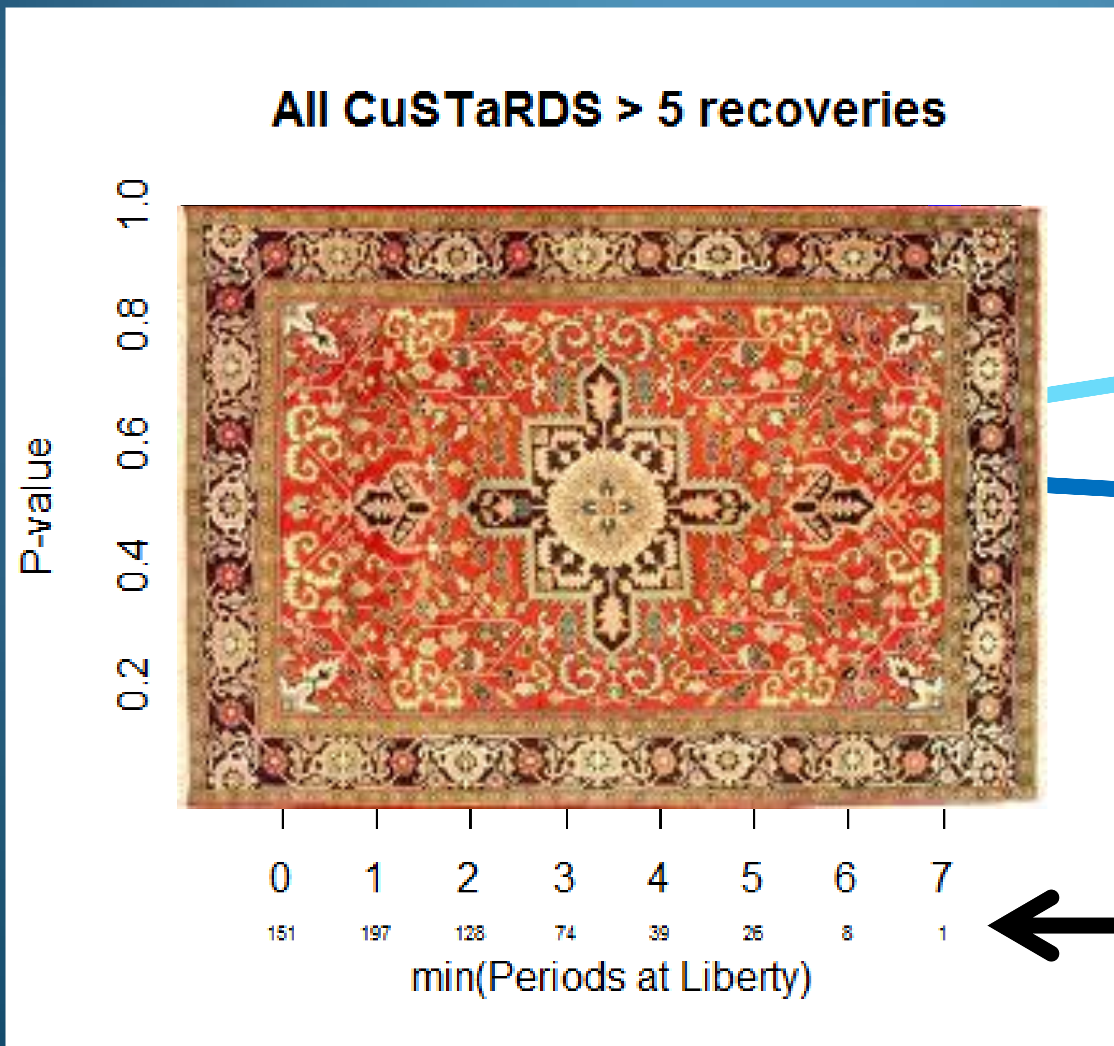
R1: 0 CuSTaRDS (Japanese Tagging Programme not yet available)

R2: 539 CuSTaRDS

R3: 19 CuSTaRDS

6 Small Release Areas

539 CuSTaRDS in SKJ Assessment Area 2 (Western WCPO)

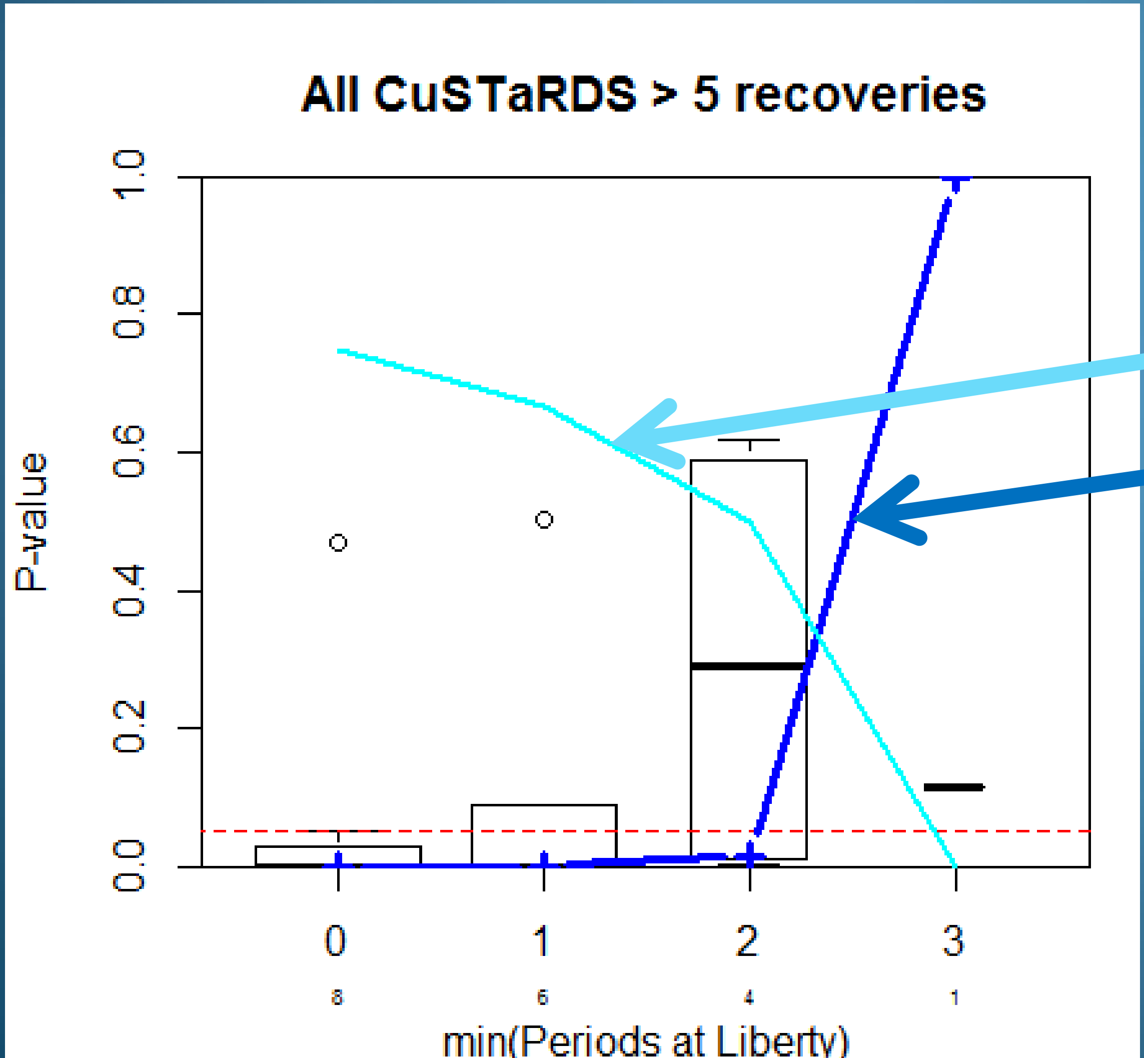


Proportion(P<0.05)

Binomial probability of obtaining this proportion (or more extreme) by chance

N (CuSTaRDS)

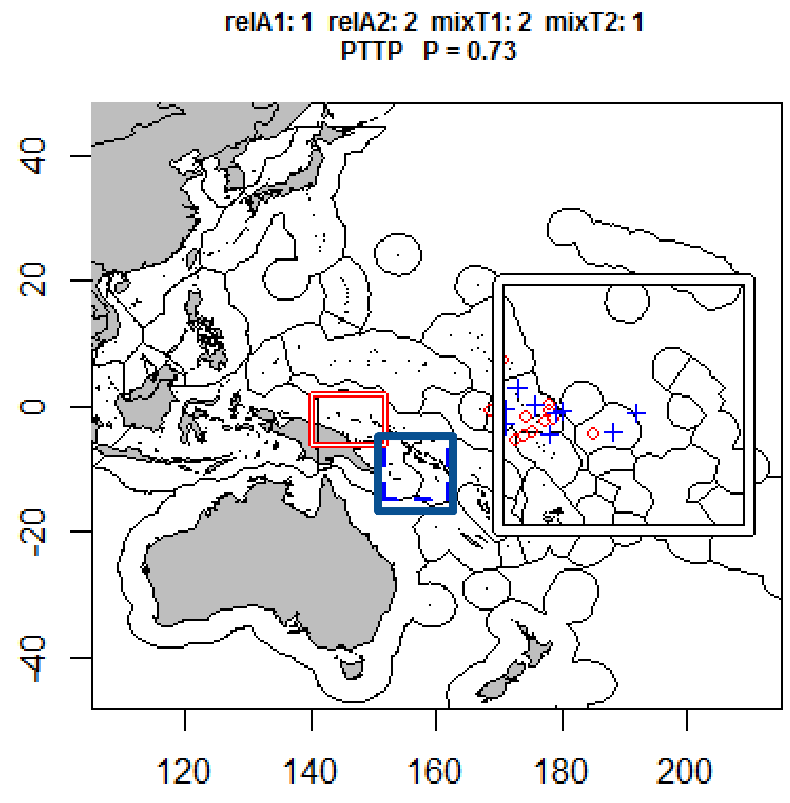
19 CuSTaRDS in SKJ Assessment Area 3 (Eastern WCPO)



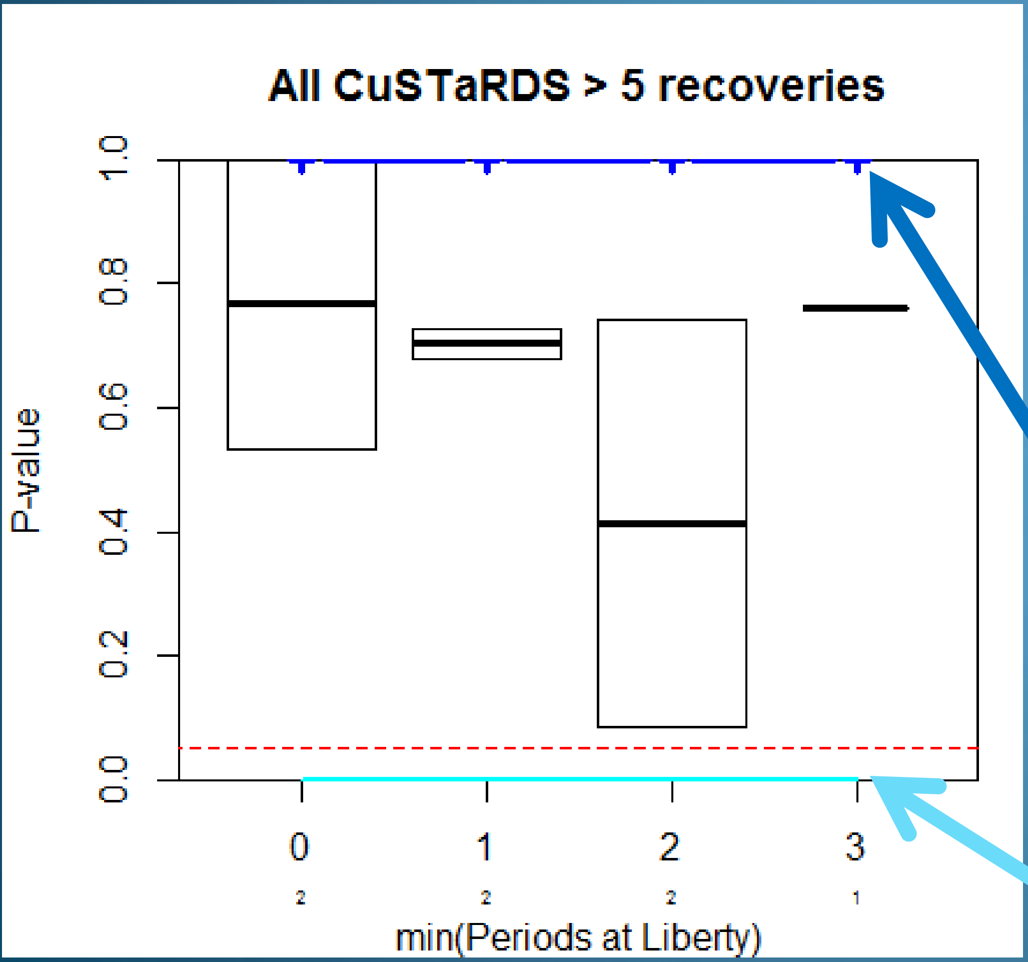
Proportion(P<0.05)

Binomial probability of obtaining this proportion (or more extreme) by chance

What about releases from SA2 recovered in SA3?



What about releases from SA2 recovered in SA3?



- Only 7 CuSTaRDs
- No evidence of incomplete mixing
- i.e. probably mixed!

Binomial probability of obtaining this proportion (or more extreme) by chance

Proportion(P<0.05)

CuSTaRD evidence for non-mixing is "Statistically Significant"

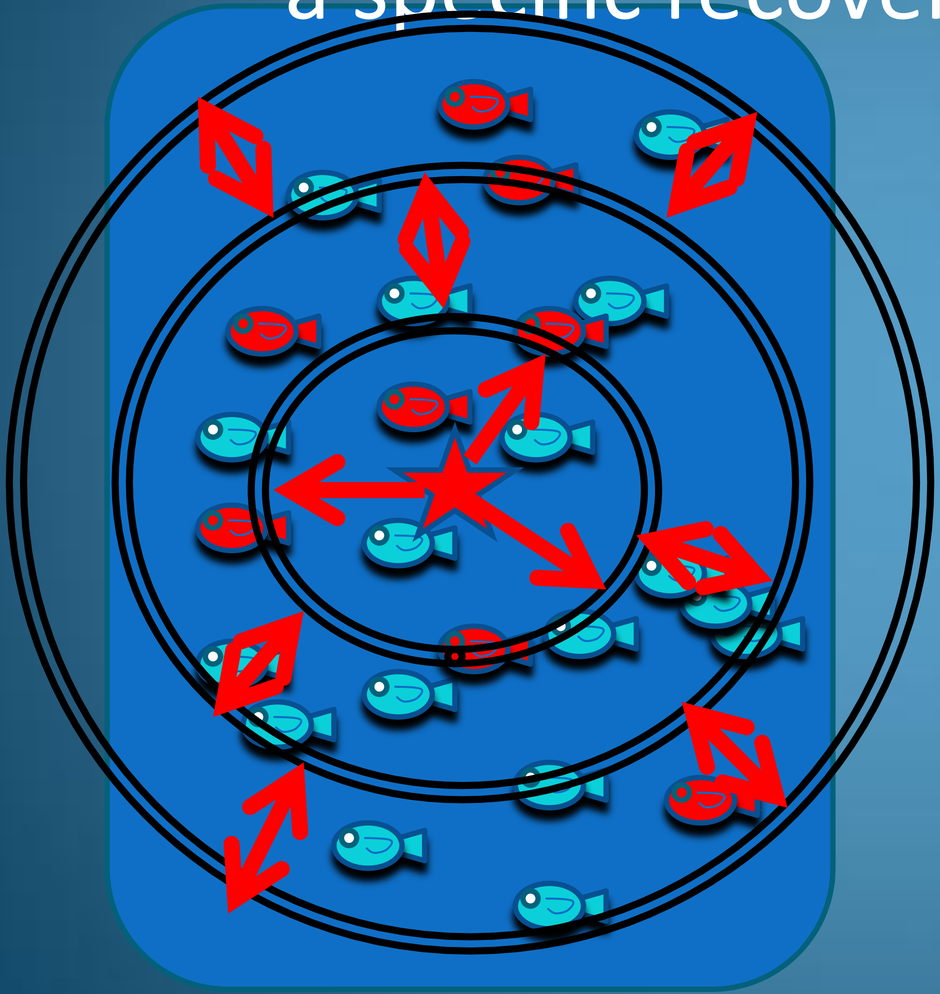
But what about Practical Signifiance ?

2) The Doughnut Analysis

Doughnut Analysis

- Similar to CuSTaRD analysis: examines tag recovery distributions from specific release events in discrete time windows
- Requires explicit recognition of catch distributions by size and tag reporting rates
- Calculates Tag density Σ(Recoveries)/Σ(Catch) as an analogue of F, for various spatial regions

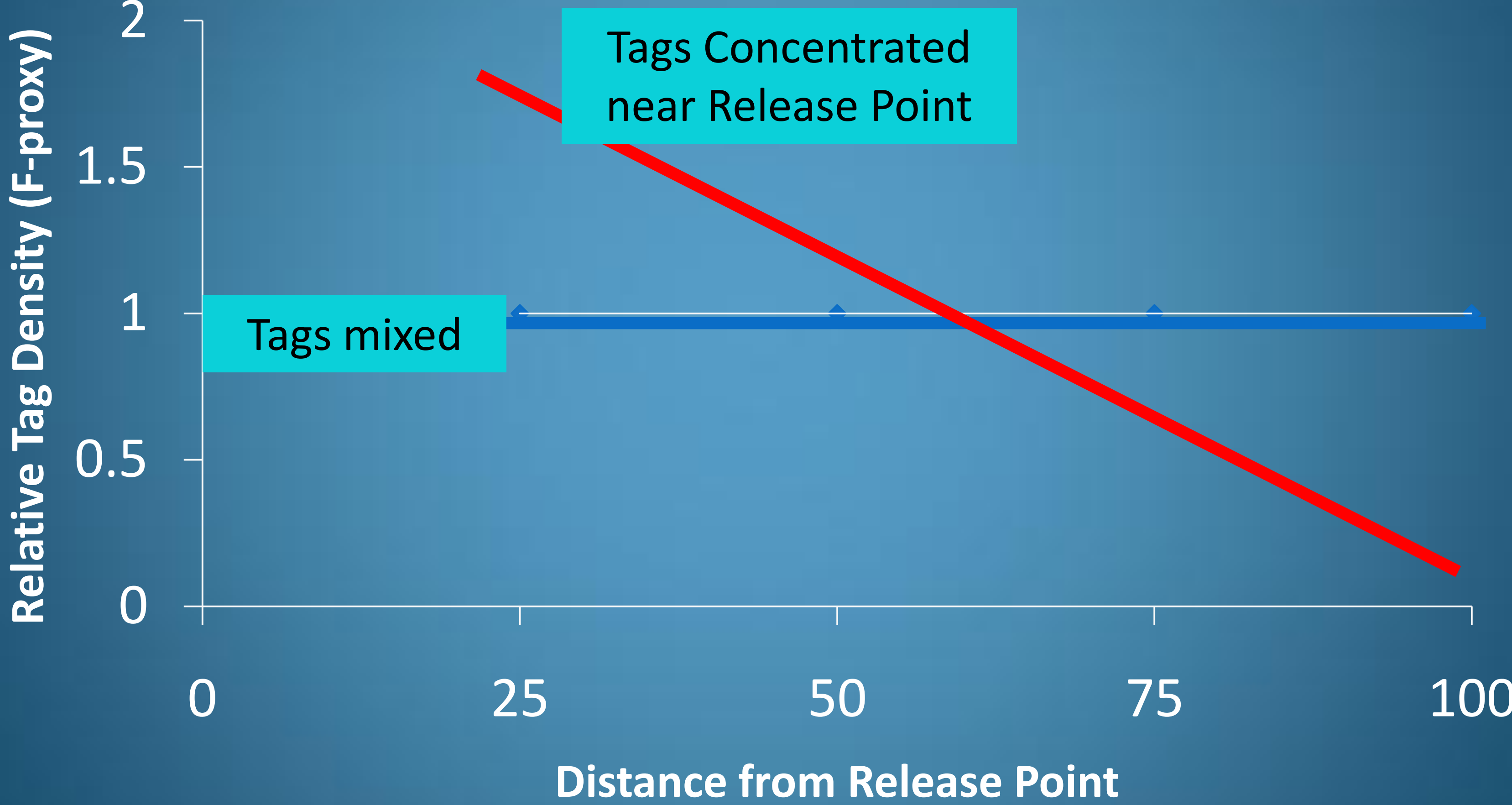
Calculate Tag Density in concentric regions (Doughnuts) for a specific release event in a specific recovery time window



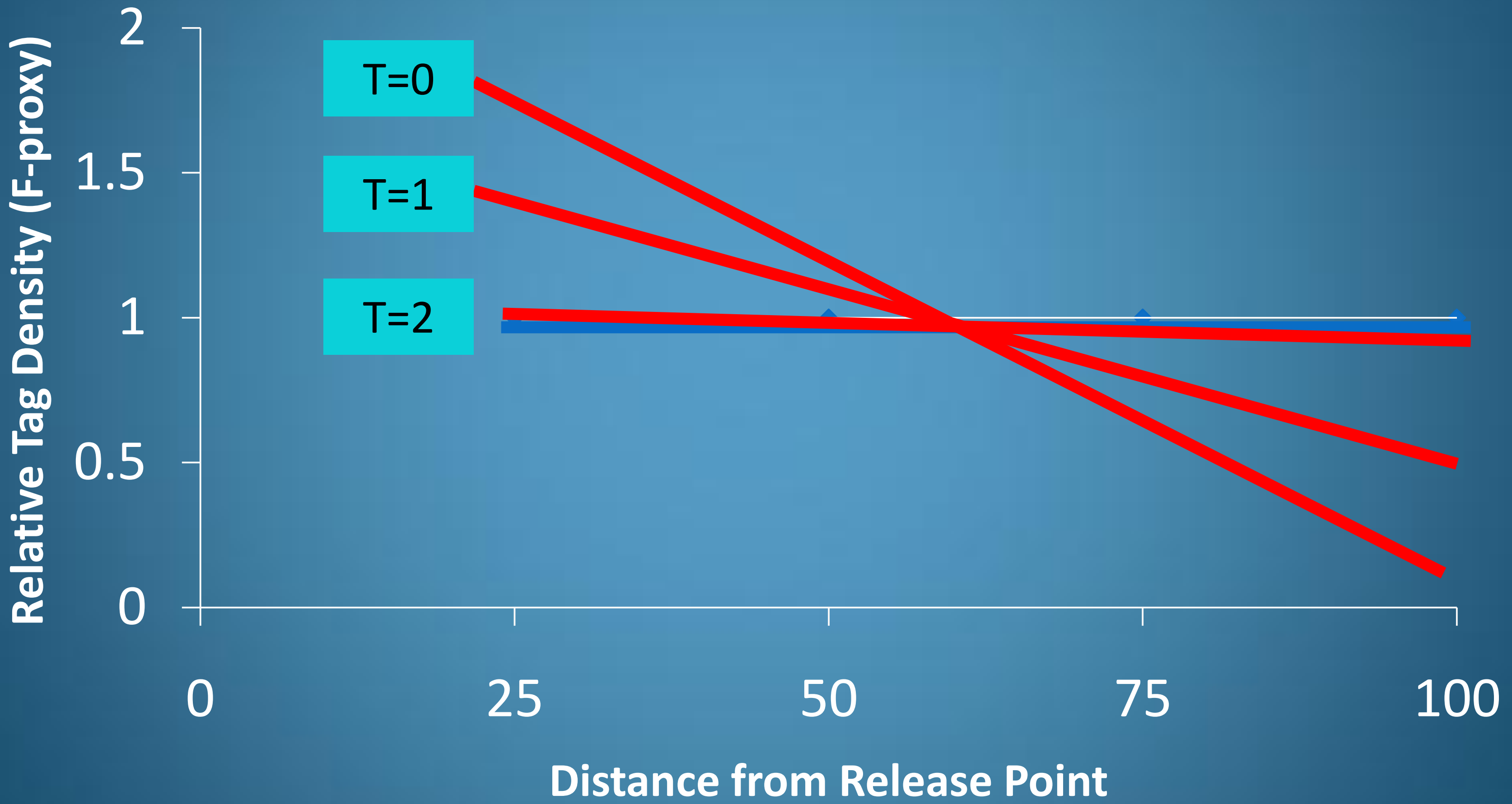
e.g. Distance from Release Points:

- 0 – 25 %ile
- 25% - 50%
- 50 – 75%
- 75-100%

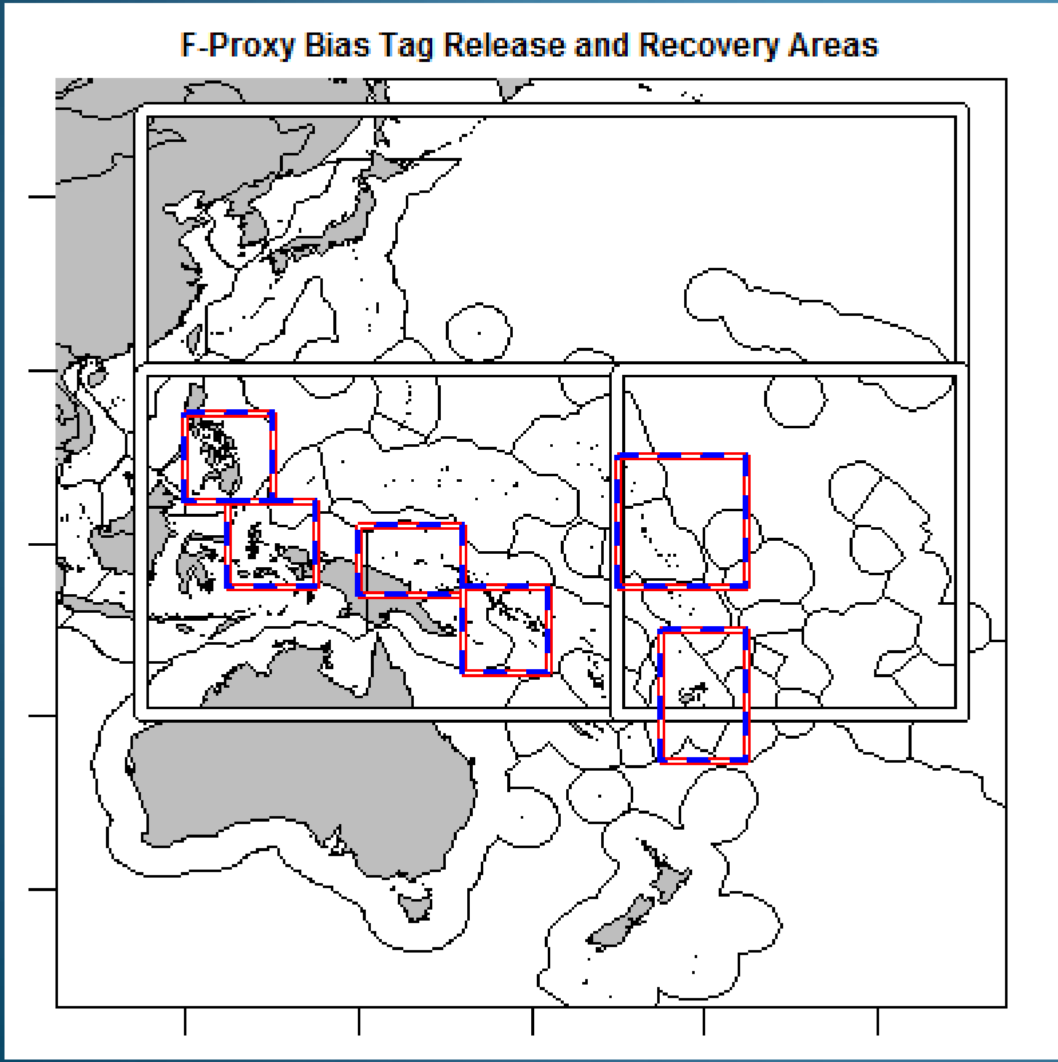
Ideally:
(Tag Density by Quartile) / (Total Tag Density)



Mixing hopefully increases over time

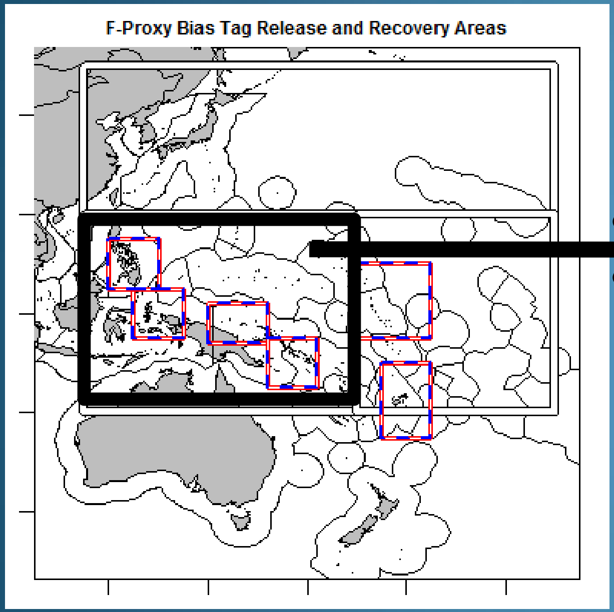


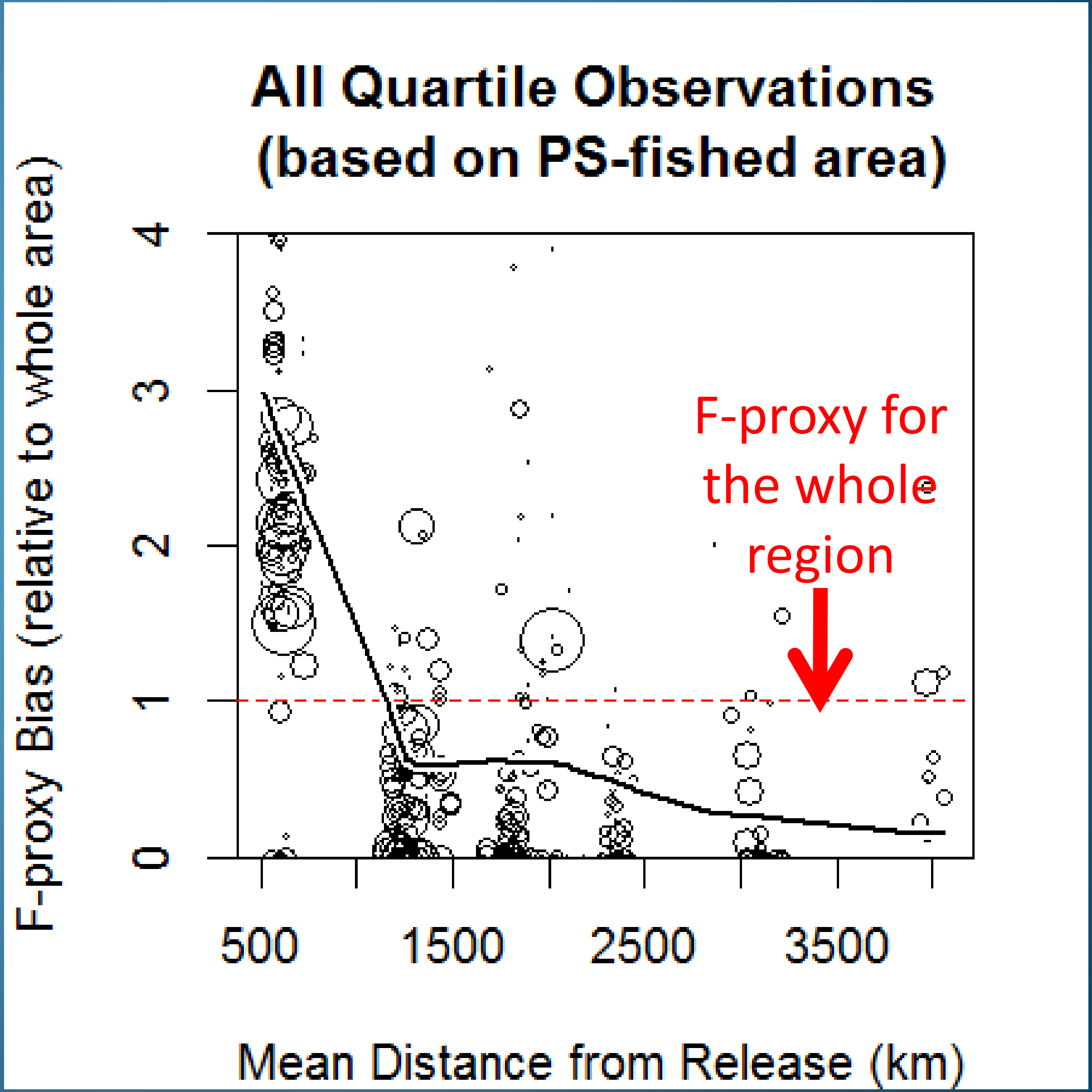
The doughnut analysis with real data



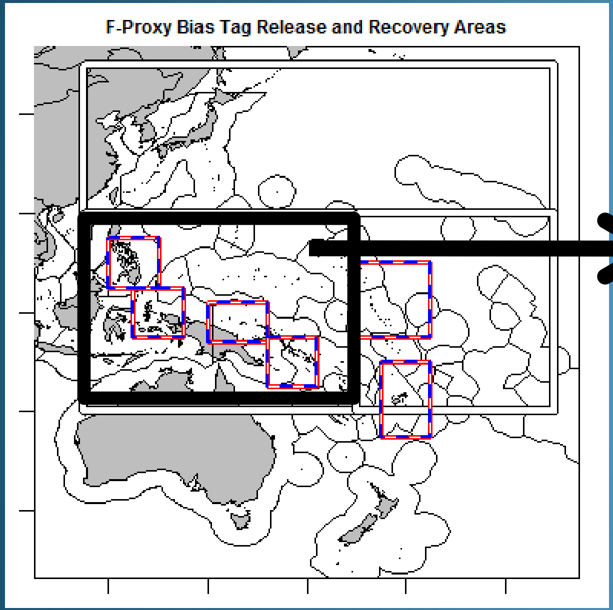
- WCPO SKJ SA region
- 6 release areas
- ~75 Doughnut events

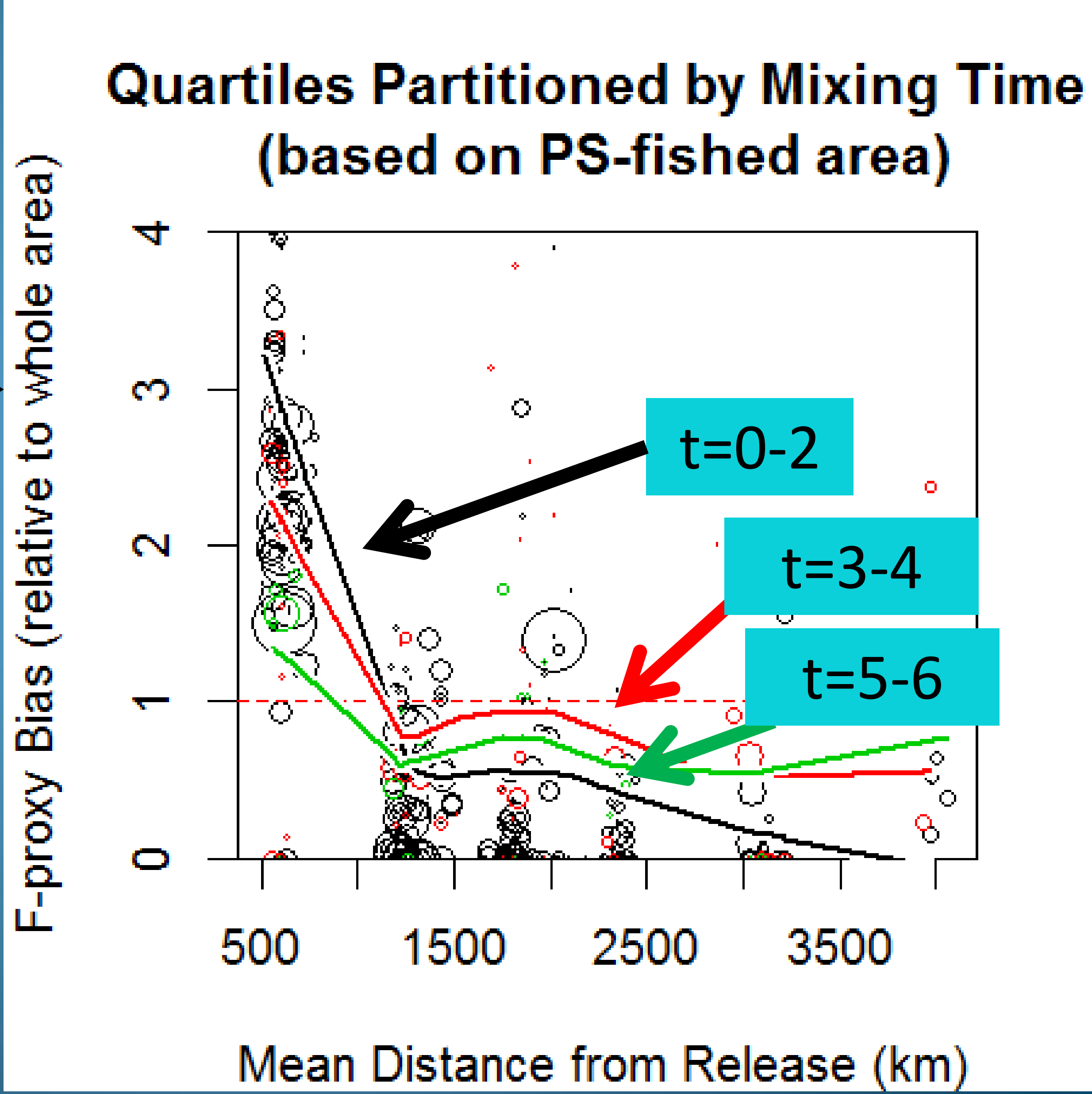
What does the F-proxy spatial bias look like for recoveries in stock assessment area 2?



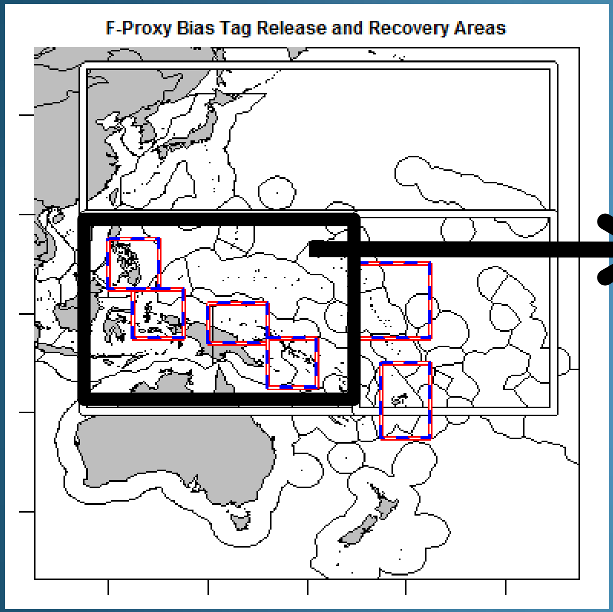


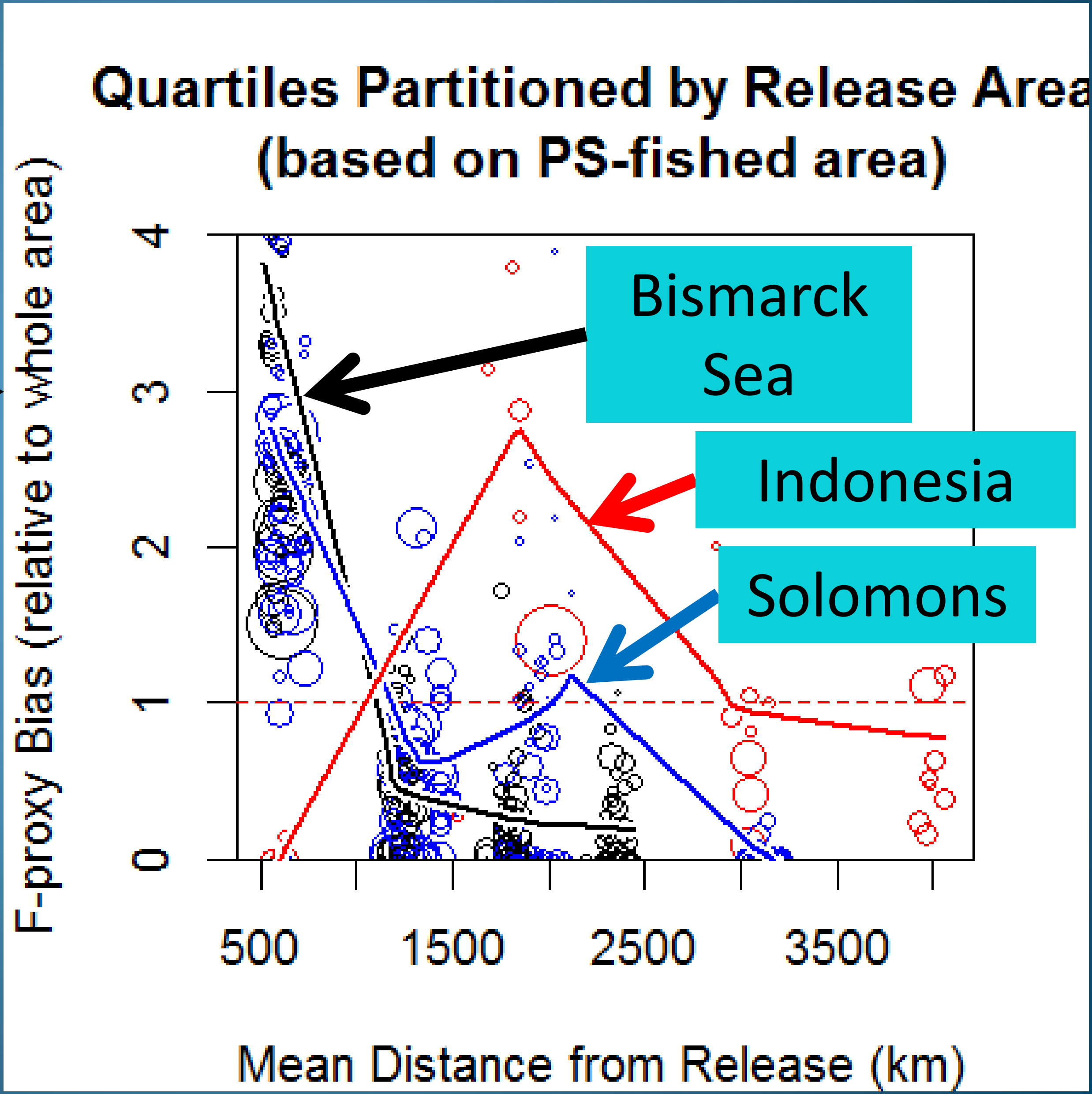
F-proxy spatial bias partitioned by mixing time





F-proxy spatial bias partitioned by Release Area





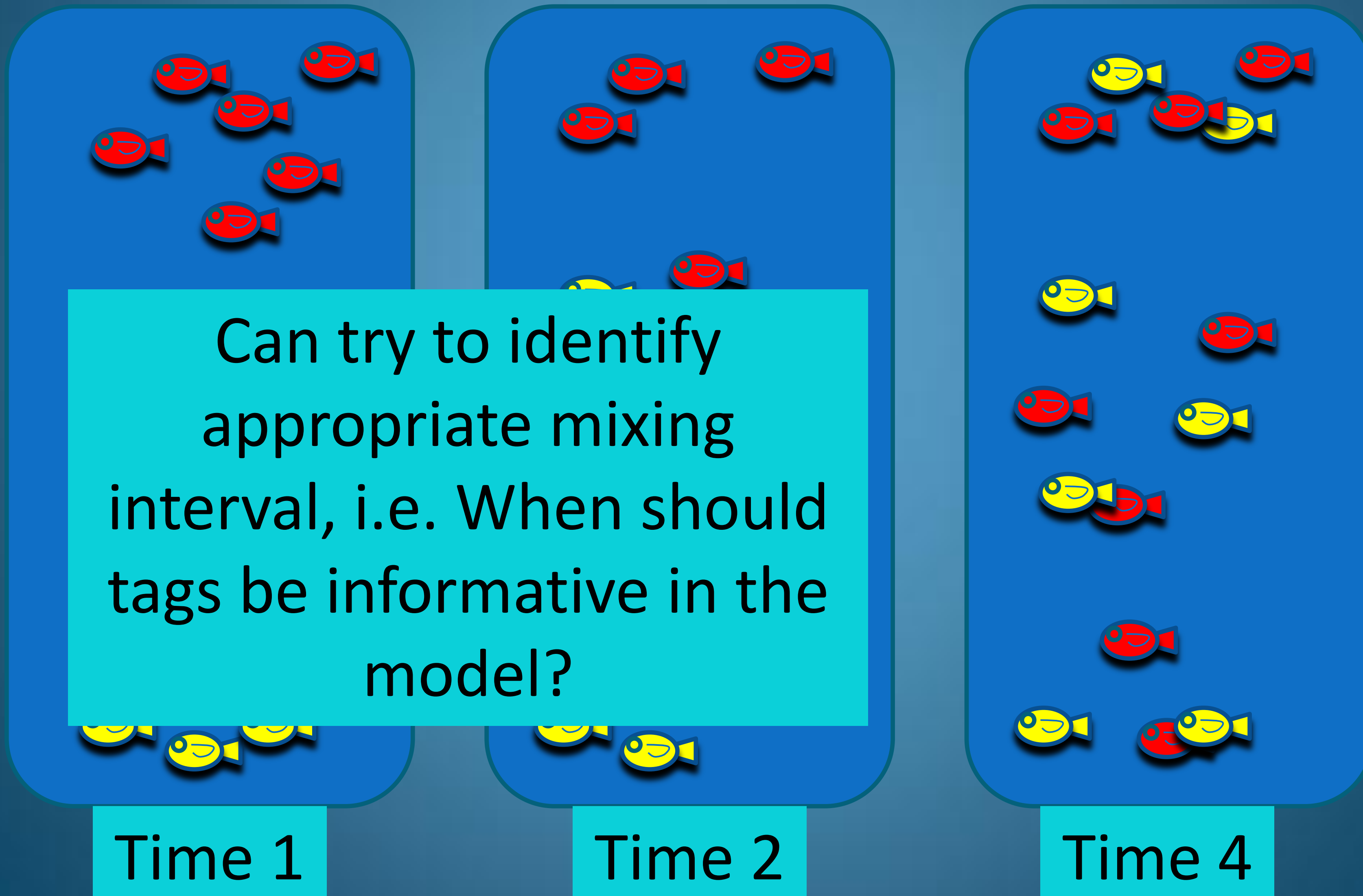
Doughnut Conclusions

- As with CuSTaRD: tags are clearly not mixing at a desirable rate for the current assessment structure.
- F estimates potentially very biased by incomplete mixing, but the bias magnitude is mitigated by an unknown extent by the broad distribution of the fishery.
- Longer mixing period seems to correspond to reduced potential F bias, but still evident after 5-6 quarters

Can these analyses be used to better formulate the assessment?

Addressing questions of model formulation:

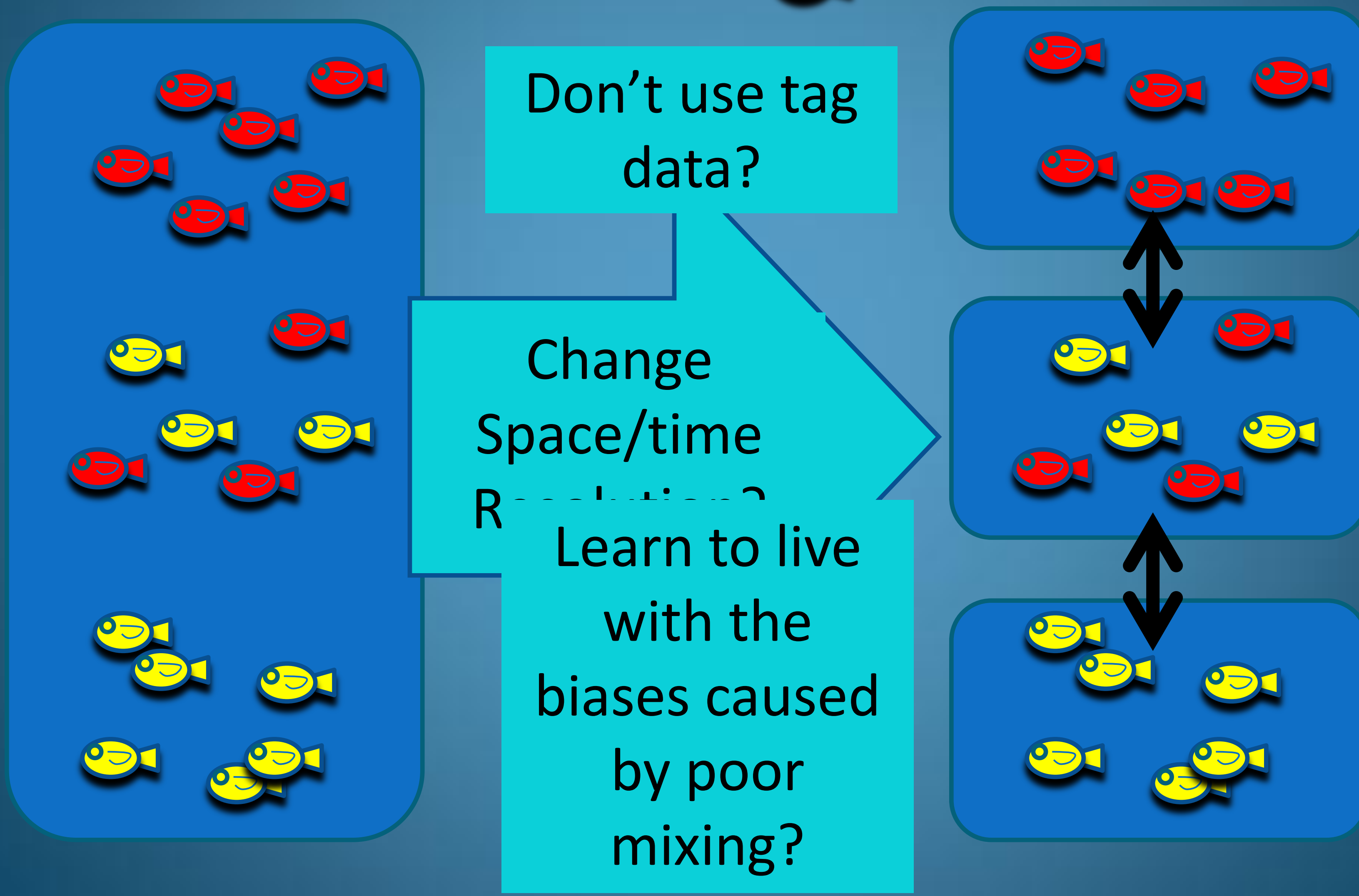
- Release Event 1
- Release Event 2



Can try to identify appropriate mixing interval, i.e. When should tags be informative in the model?

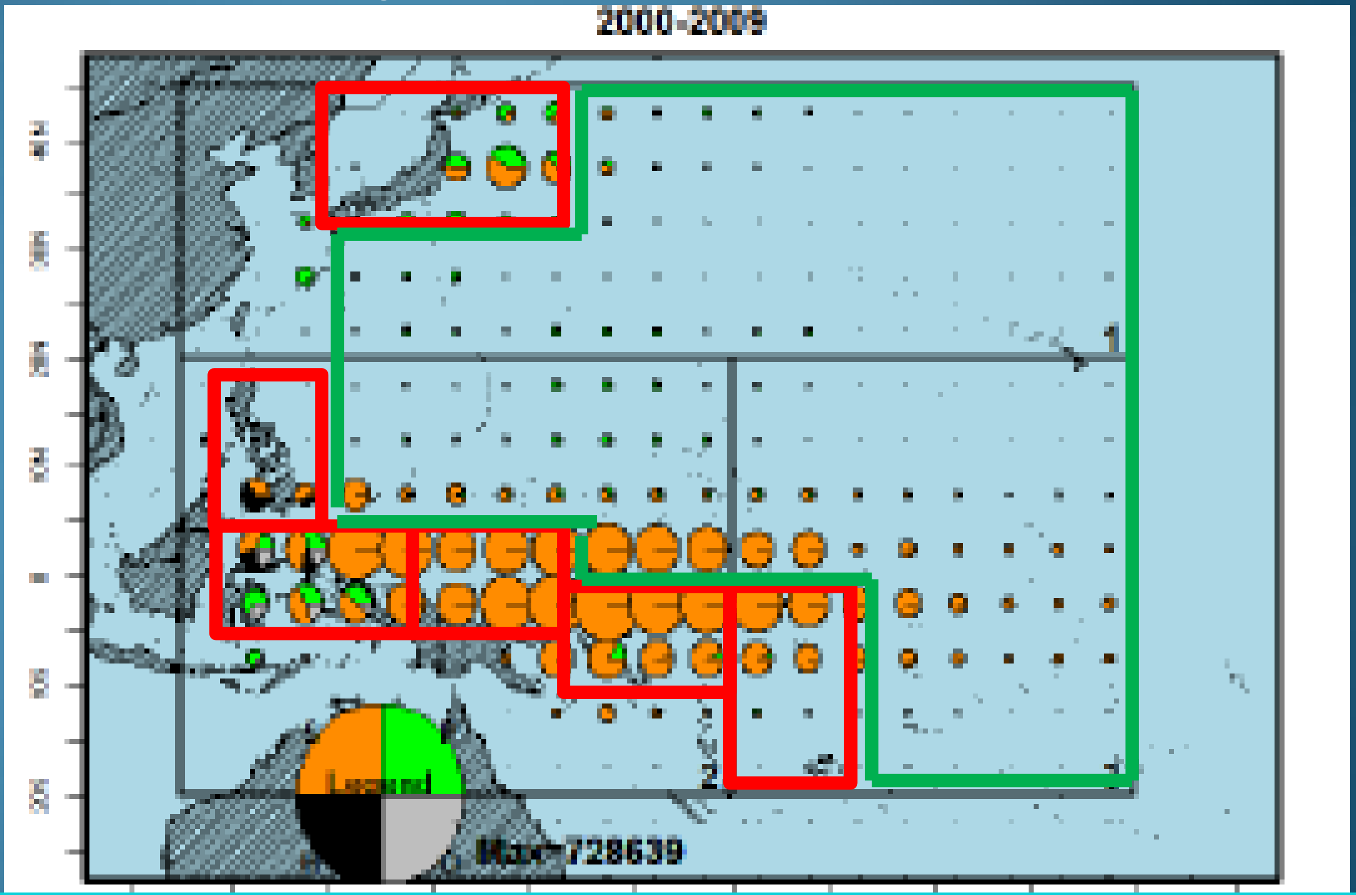
What if mixing is too slow or never occurs?

- Release Event 1
- Release Event 2



Toward a structure that meets mixing assumptions....?

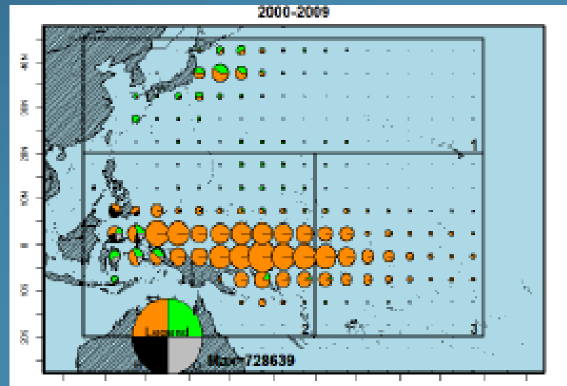
- Small coastal units
- Large Pelagic units

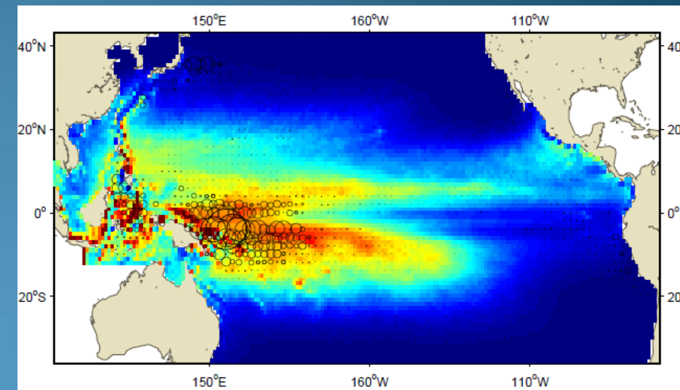


Find a grid structure in which:

1) Can no longer detect evidence of non-mixing (due to real mixing or low power)
2) Or: F-proxy bias small enough to ignore
3) Or: Can develop crude bias-corrections?

But mixing is not everything...





Aggregated Models

Current MFCL Assessment

e.g. SEAPODYM,TAGEST

Decreasing Aggregation Bias

Simulations of estimator performance
in management context ?

Increasing Estimation Variance
Greater Data and Computational Requirements.
Hobbled by tag release design and recovery errors?

Are we asking the right questions?

Conclusions I

- CuSTaRD and Doughnut analyses are simple tools that can identify situations when tags are not meeting standard mixing assumptions.
  - CuSTaRD requires less data
  - Doughnut works in units analogous to fishing mortality

- Limitations:
  - More powerful with more release areas and more recoveries
  - Can never prove that full mixing has occurred
  - Do not really distinguish between "statistical significance" and practical implications

Conclusions II: WCPO SKJ Assessment

- Mixing assumptions are not currently being met
- Supports prevailing biological notions:
  - Residency in archipelagic waters
  - Rapid displacements in open water
- Suggests a direction for re-structuring the assessment to reduce the impact of mixing assumption

Many thanks to:

- the SPC and the good people of PNG for funding the work.
- John Hampton, Sylvain Caillot, Simon Nicol, Emmanuel Schneiter and Caroline Sanchez for providing helpful explanations of the data and fisheries
- Persons not yet identified who will hopefully find the embarrassing mistakes before they are published

Images Courtesy of FreeDigitalPhotos.net

Comparison of analyses

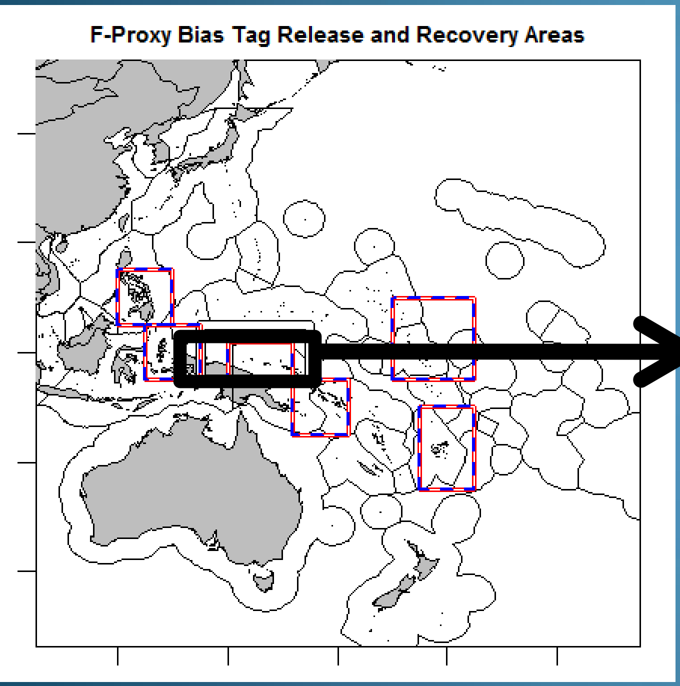
CuSTaRD

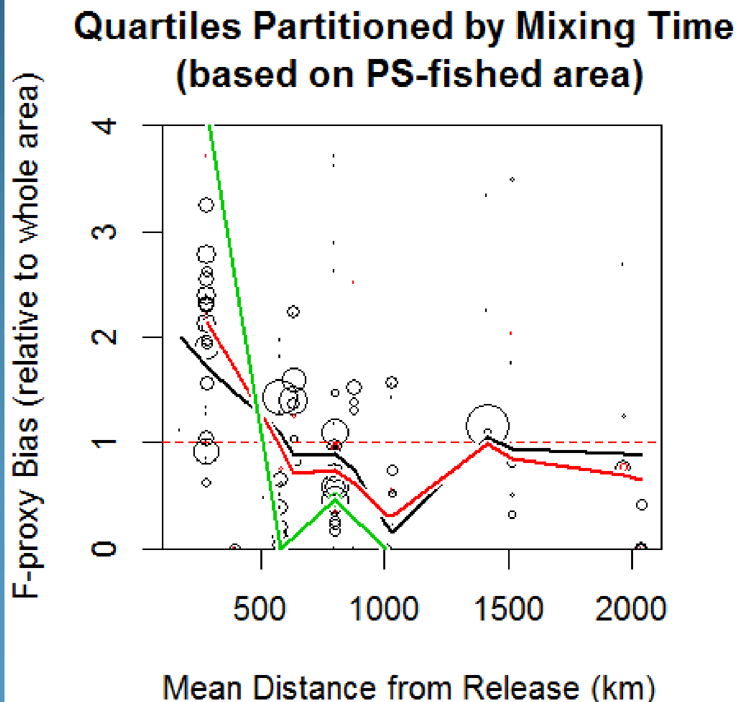
- Does not require catch or catch-at-size
- Can use recoveries from all fleets
- Qualitative – can detect non-mixing, but does not tell us anything about the implications
- Low power, but powerful enough to find mixing problems in WCPO

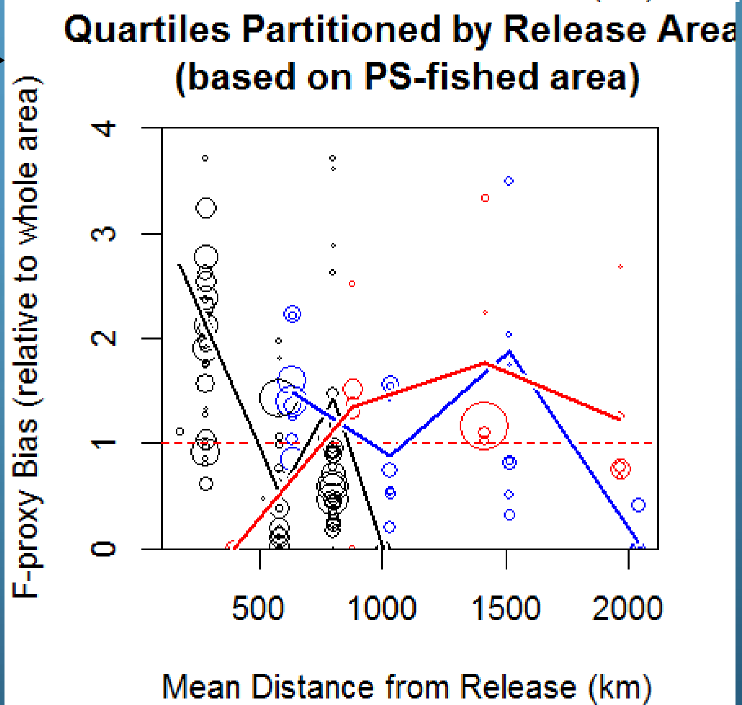
Doughnut

- Requires catch-at-size information
- Need to be careful pooling fleets with unknown reporting rates
- Quantitative – identifies potential biases in something analogous to F
- Should be higher power
- More sensitive to tag recovery position errors

How does the F-proxy bias change within smaller recovery areas?

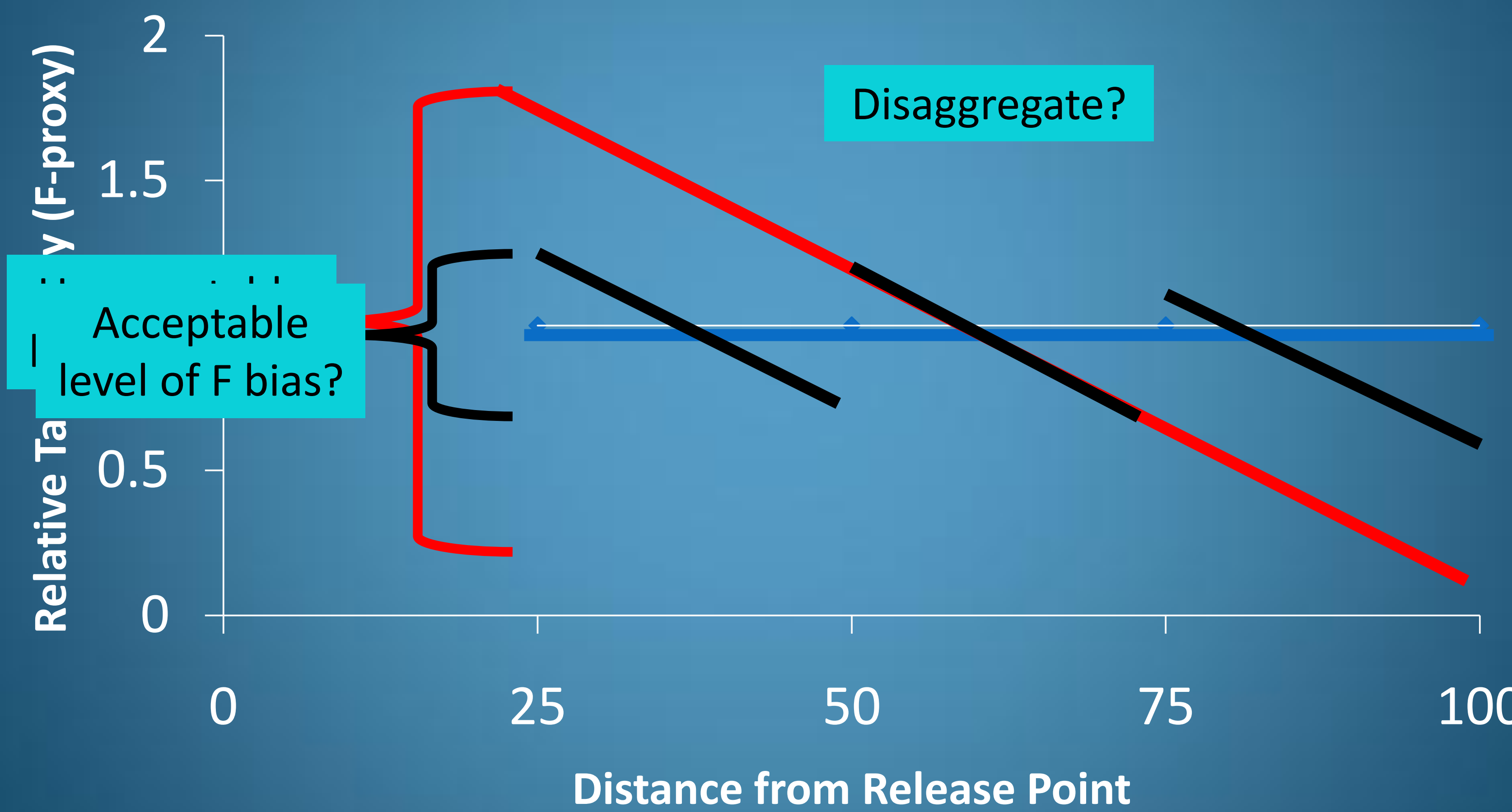




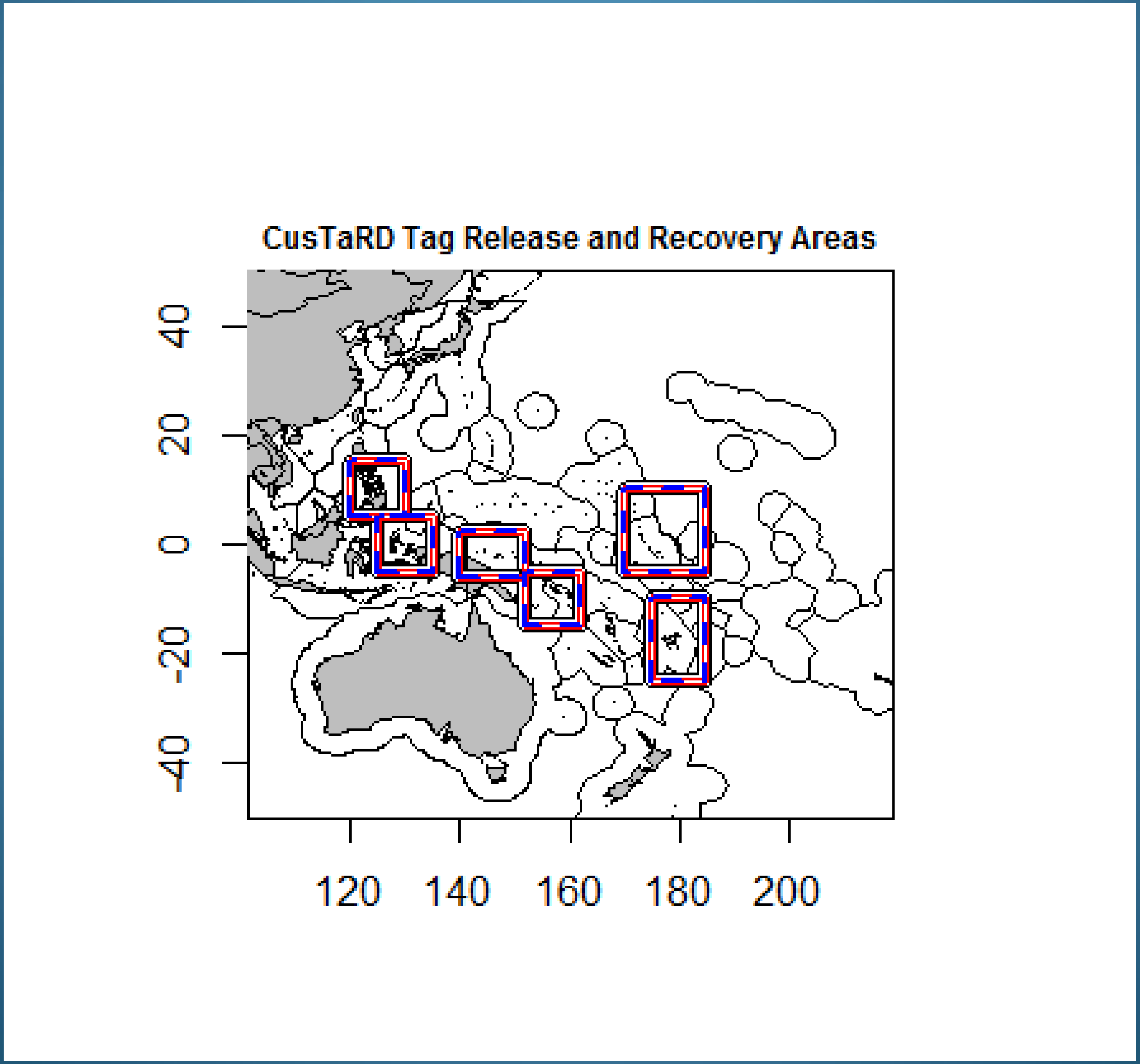


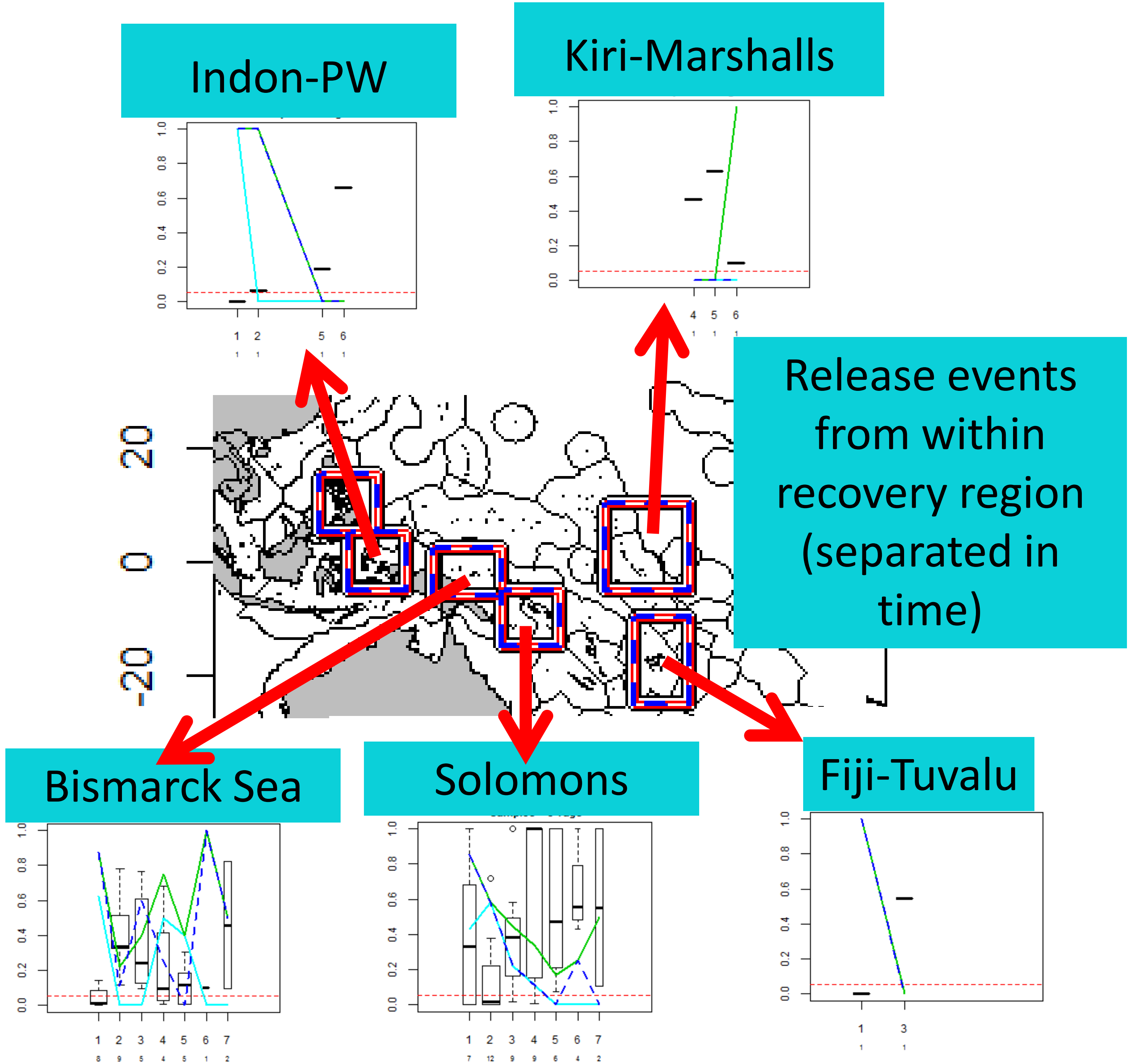
- Pattern of F bias with mixing time and distance from release might decrease
- Signal to noise ratio is also reduced

If Mixing is too slow



What does mixing look like within the small release areas?





Indon-PW
Kiri-Marshalls
Release events from within recovery region (separated in time)
20
0
-20
Bismarck Sea
Solomons
Fiji-Tuvalu

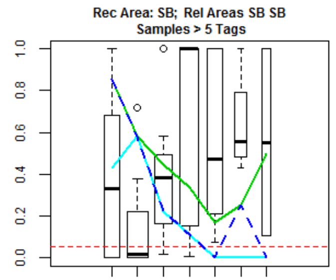
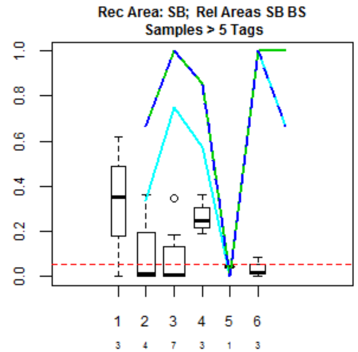
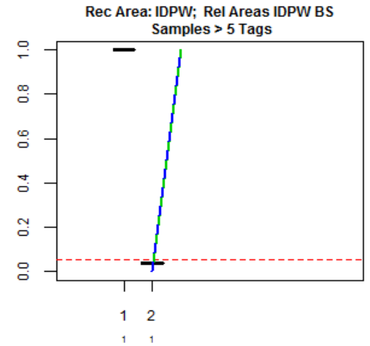
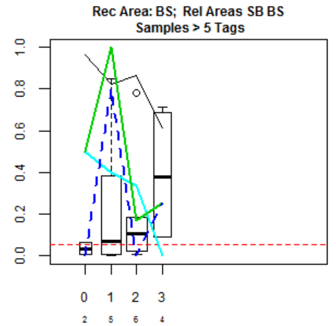
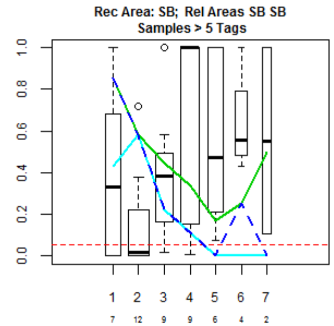
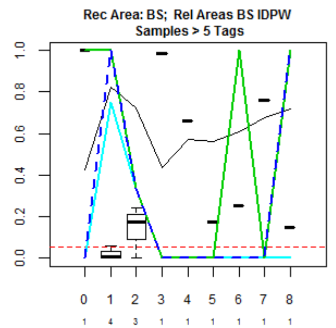
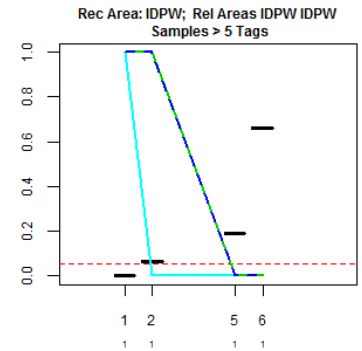
Ideally would like to look at mixing rates among all regions

	Philippines	Indonesia-Palau	Bismarck Sea	Solomon Islands	Marshalls-Kiribati	Fiji-Tuvalu
Philippines						
Indonesia-Palau						
Bismarck Sea						
Solomon Islands						
Marshalls-Kiribati						
Fiji-Tuvalu						

However, not enough observations outside of Bismarck Sea and Solomon Islands

Recovery region and release region 1

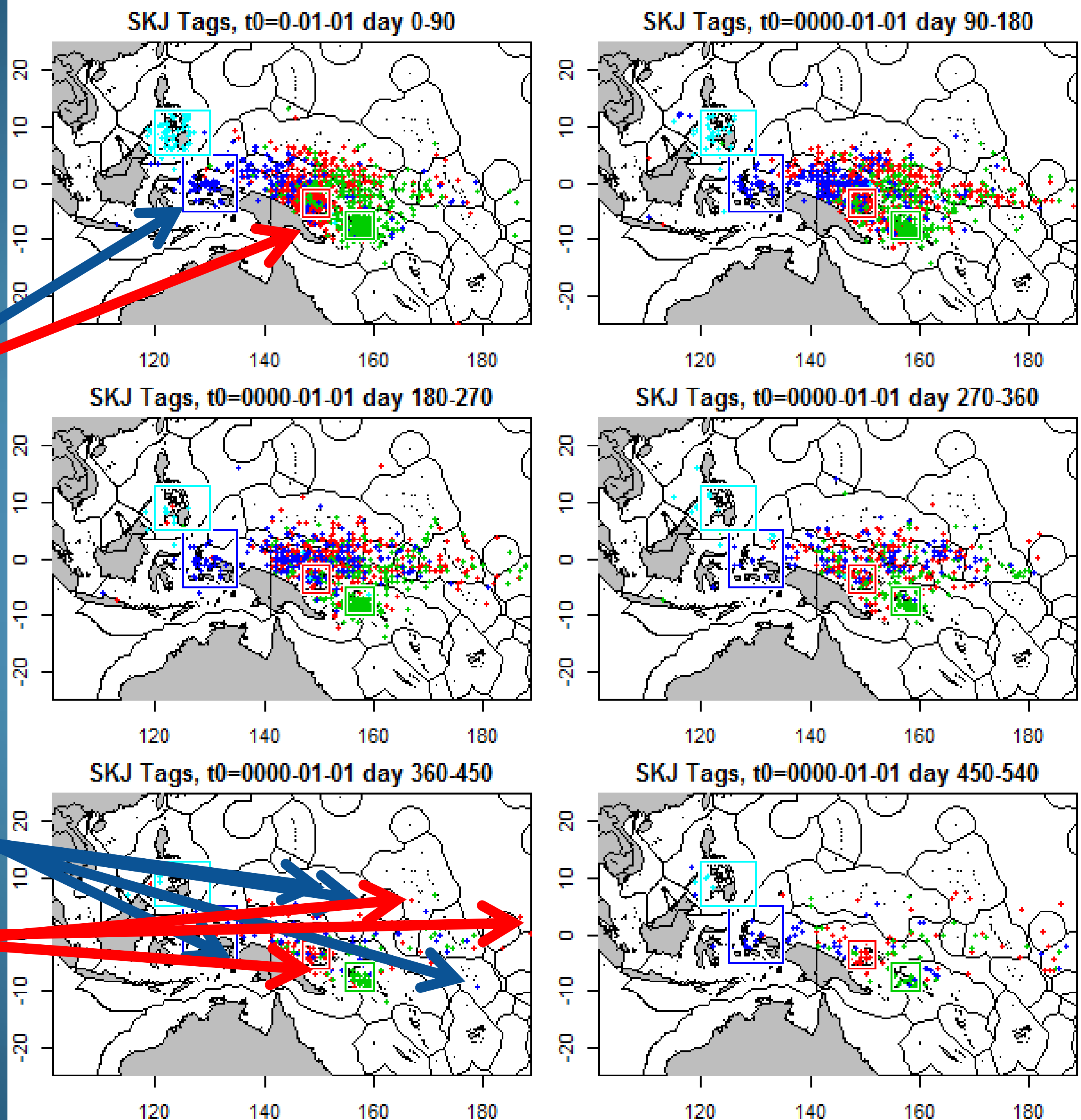
Release region 2

	Bismarck Sea	Solomons	Indonesia-PW
Bismarck Sea			
Solomons Islands			No observations
Indonesia PW		No observations	

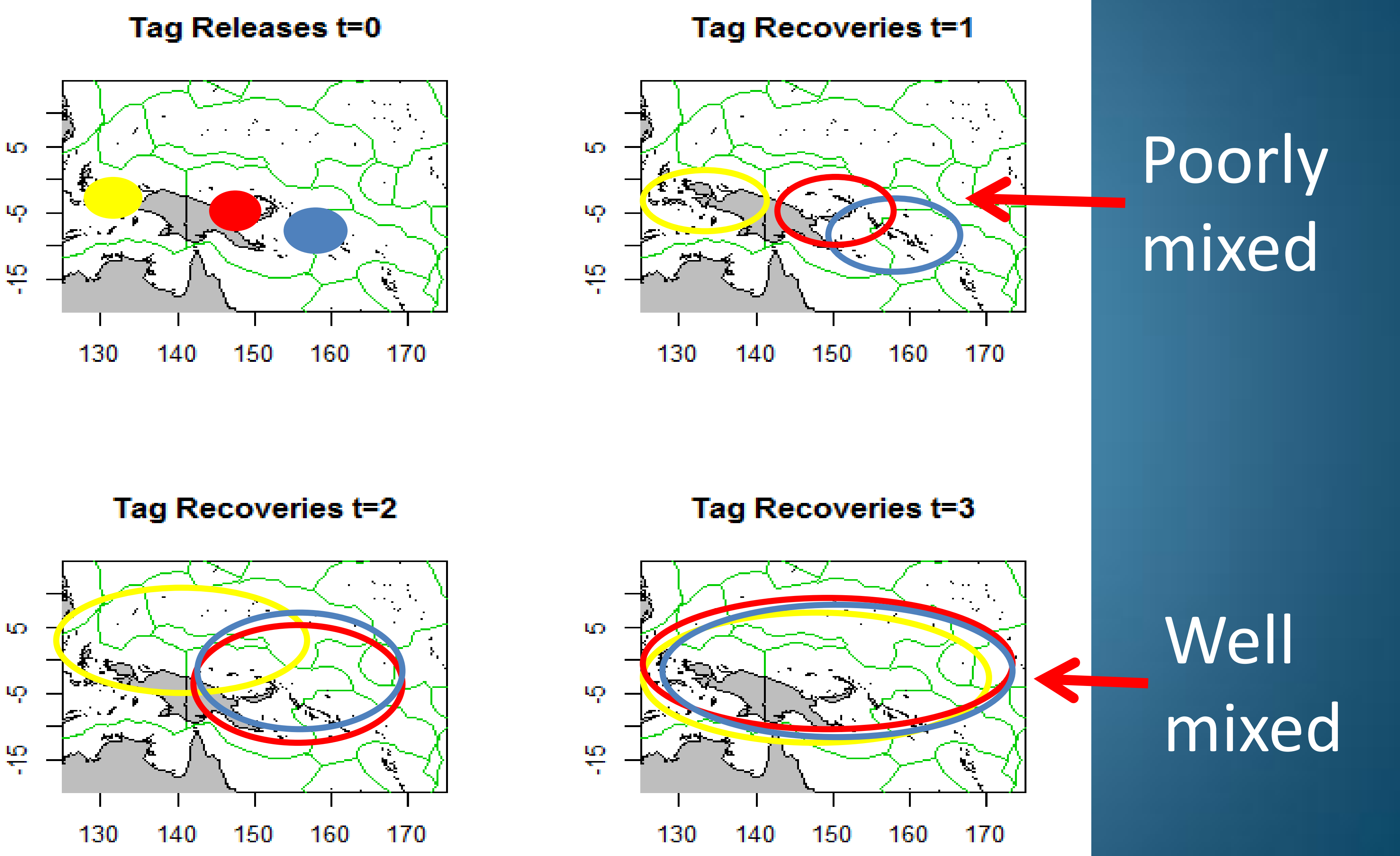
Extra bits too good to throw away yet...

Square release areas

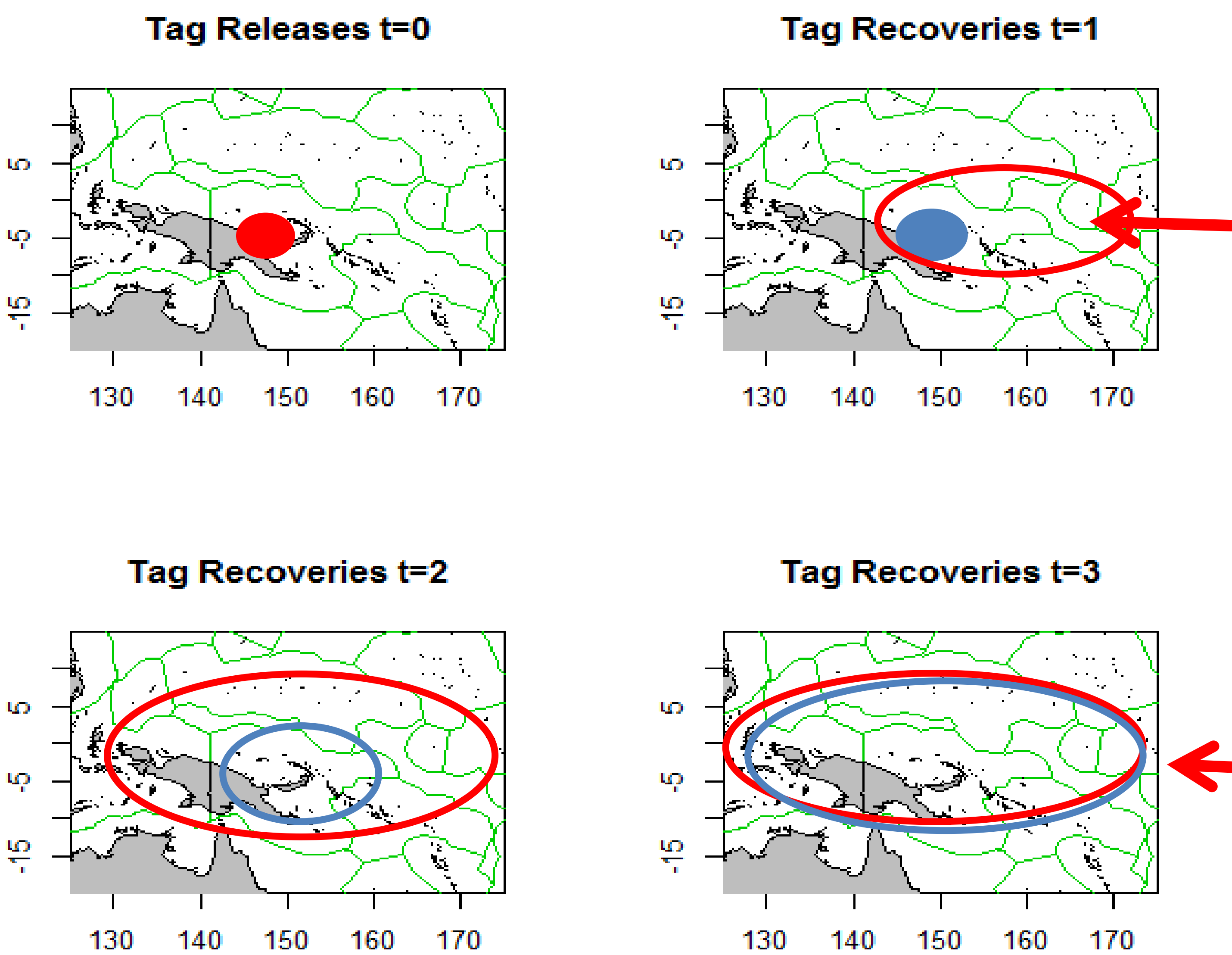
Points are Recoveries



e.g. 3 Spatially separated release events



e.g. 2 Temporally separated release events



Poorly mixed

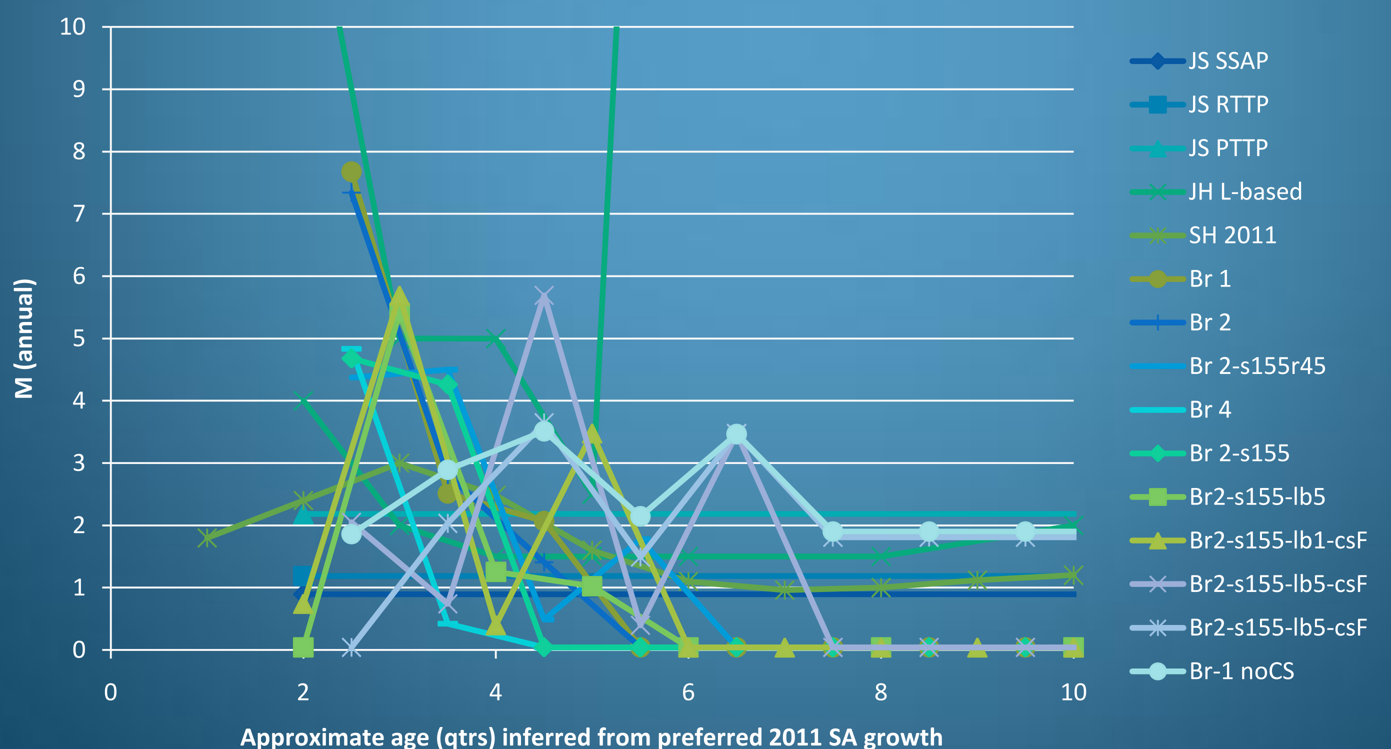
Well mixed

Practical significance?

- Comparison of spatial structures using Brownie estimators for M(a),F(t,a) and movement()
- Derived from Eveson, Laslett and Polacheck approach and software, with some changes:
  - Implicit method for stable movement estimates
  - Accounting for release timing of tags within time period
  - Lengths assigned to multiple ages to account for variability in length-at-age
- Differences from Multifan-CL
  - Fast
  - Potentially Greater freedom from stationarity assumptions
  - Inferences limited to tagged cohorts

Sneak preview of Brownie M Results...

- Cannot easily compare Fs
- Can only compare Ns if/when Petersen element is added



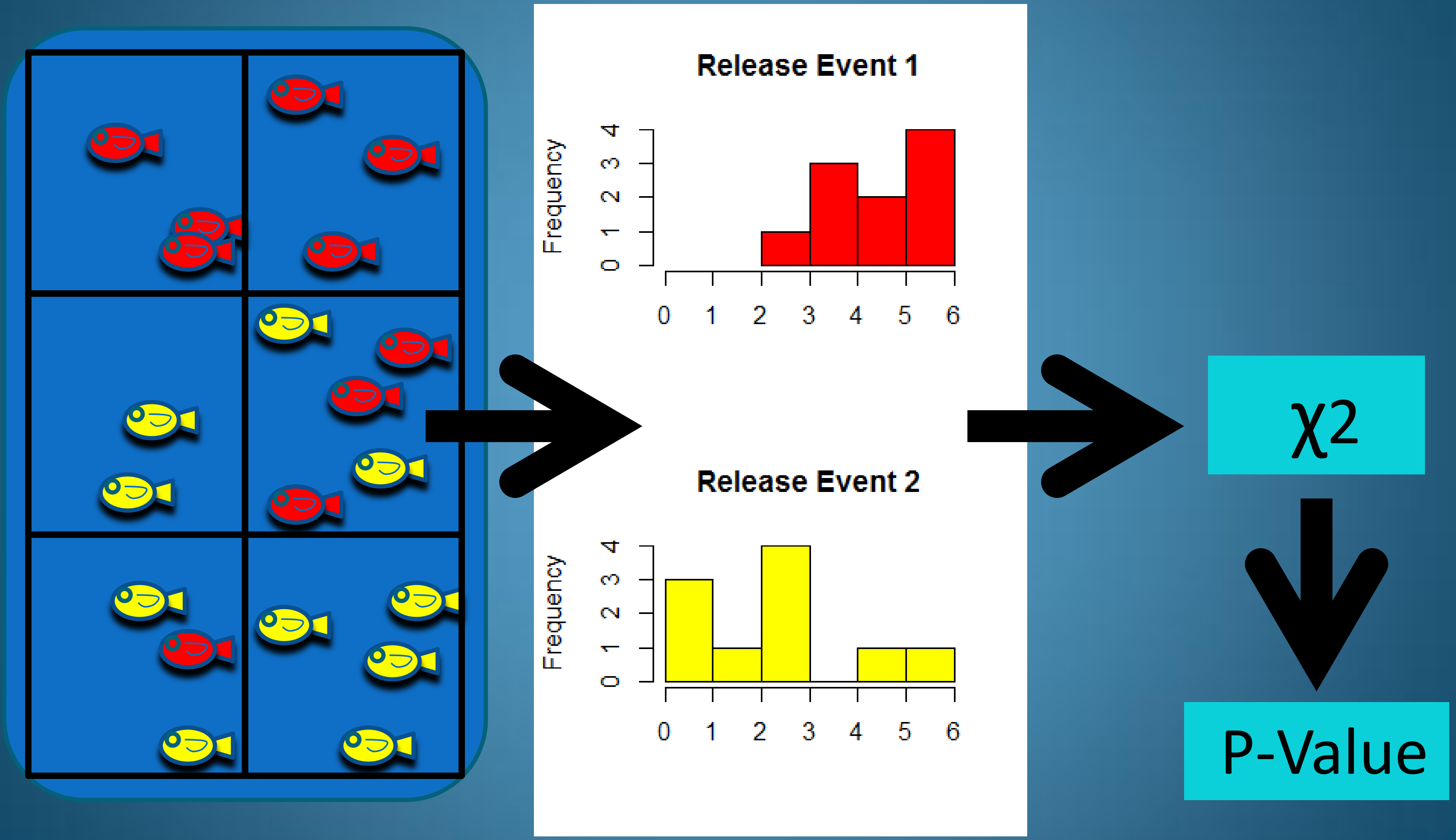
Future Work

- SKJ Assessment:
  - Continued comparison of Multifan-CL and Brownie estimators in relation to spatial (and other) assumptions
  - Increase Multifan-CL resolution?
  - Alternative modelling frameworks for spatial management questions?
- Application to YFT, BET

1) Chi-square Contingency Table on 2D grid

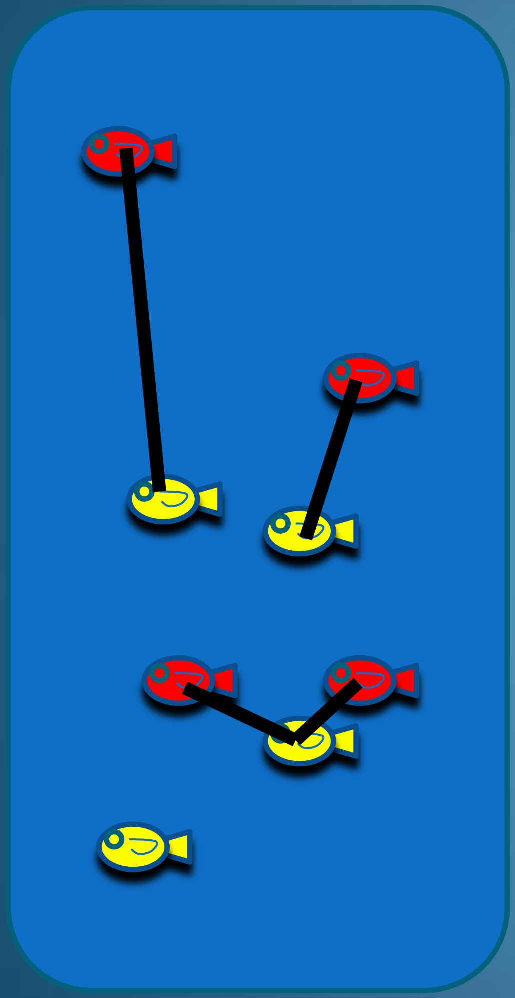
- Accounts for sample sizes
- Simple and Fast
- Appears to work in simulations
  - Individual results can be sensitive to binning assumptions
- Still doesn't sound complicated enough to be good science

Chi-square on 2D grid



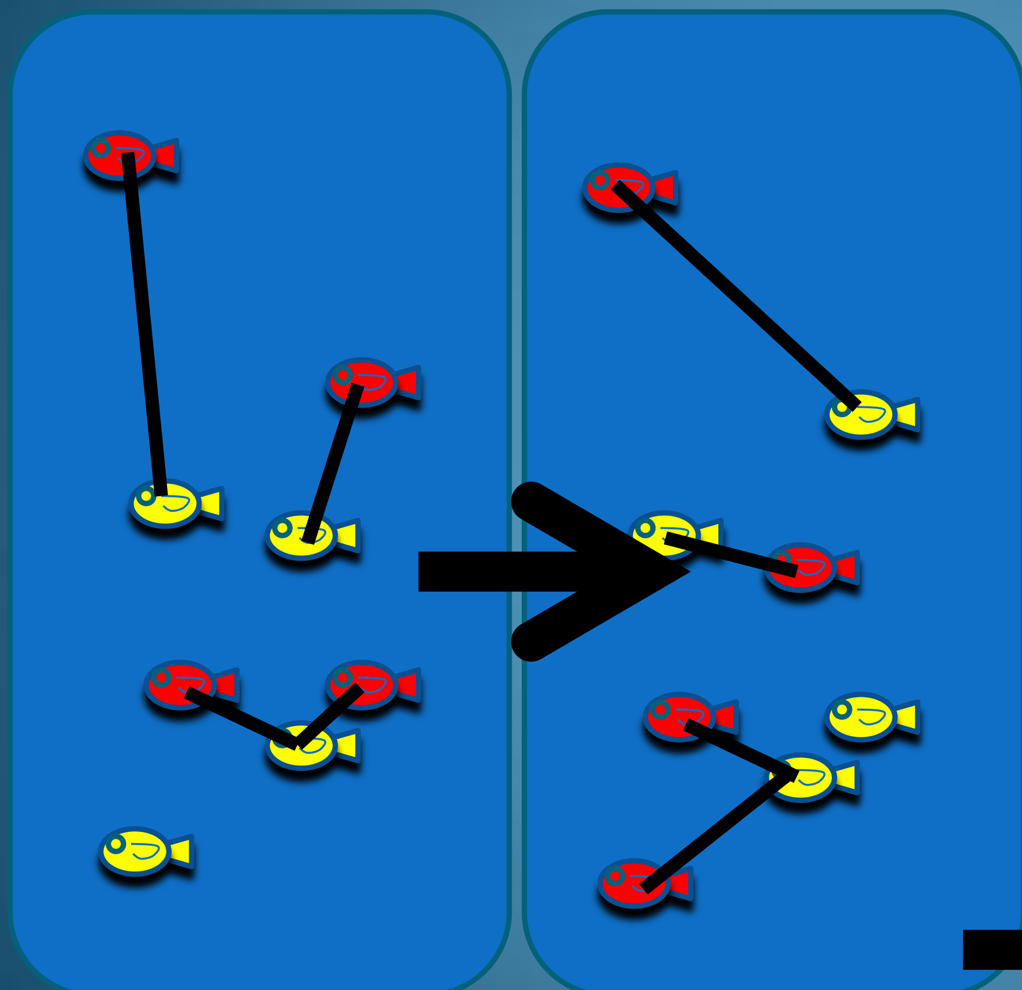
CuSTaRDs with at least 5 tags in each recovery distribution included

Nearest-Neighbour Permutation test



1) Calculate mean nearest neighbour cross:
   – Mean distance between each recovery from release event 1 to the nearest tag from release event 2.

Nearest-Neighbour Permutation test



2) Monte Carlo simulations to generate a null distribution:

- Randomly reassign each tag recovery to one of the two release events and recalculate the mean(NN).

Repeat

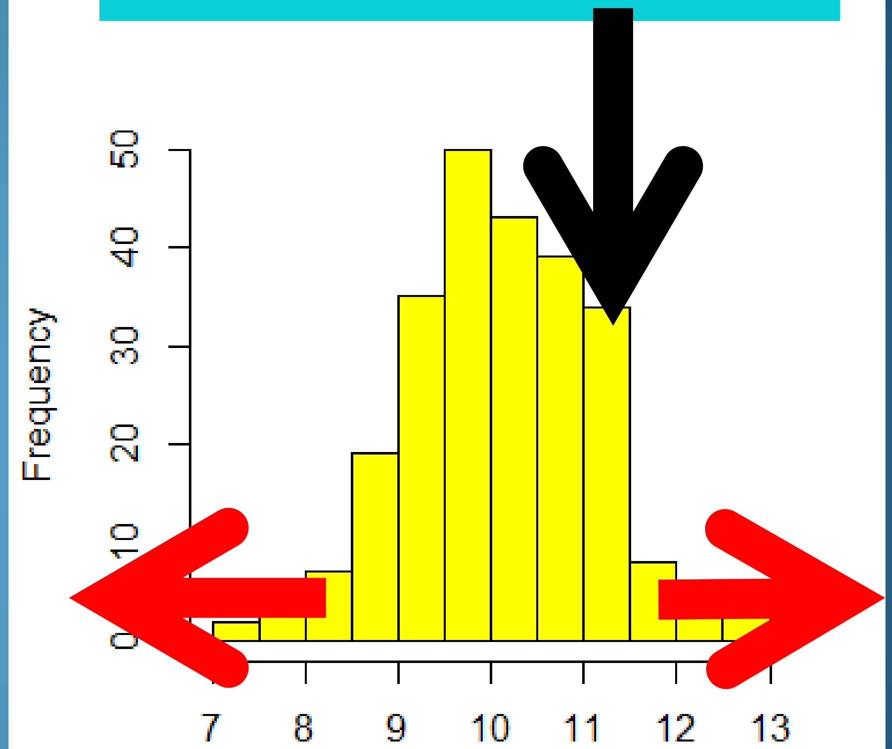
Nearest-Neighbour Permutation Test

- Null distribution from Monte Carlo simulations

- Calculate P-value by comparing observed mean(NN) with null distribution

- P-value (Index of Similarity) represents Probability of observing this degree of difference (or more) by chance, if the two distributions are actually samples from a single distribution

Observed mean(NN)

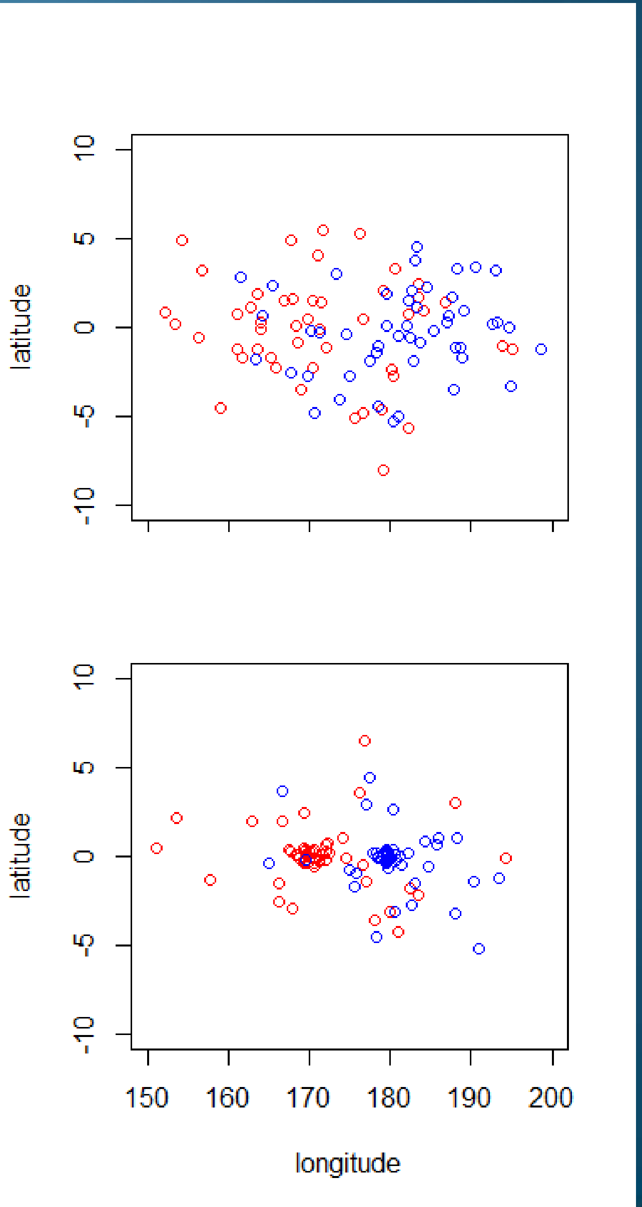


$p<0.025$ or $p>0.975$

Strong evidence that distributions are different

Simulations of Chi-square and NN-P Indices of Spatial similarity

- Paired tag recovery distributions:
  - Sample sizes of 5-80 tags
  - A) Bivariate Normal Distributions
    - σ(oceanic) = 10° lon; 2.5° lat
    - Δ lon = 0 - 30 °
  - B) Bivariate Normal with 50% contamination from a second (clumped) distribution
    - σ(coast)=σ(oceanic)/10



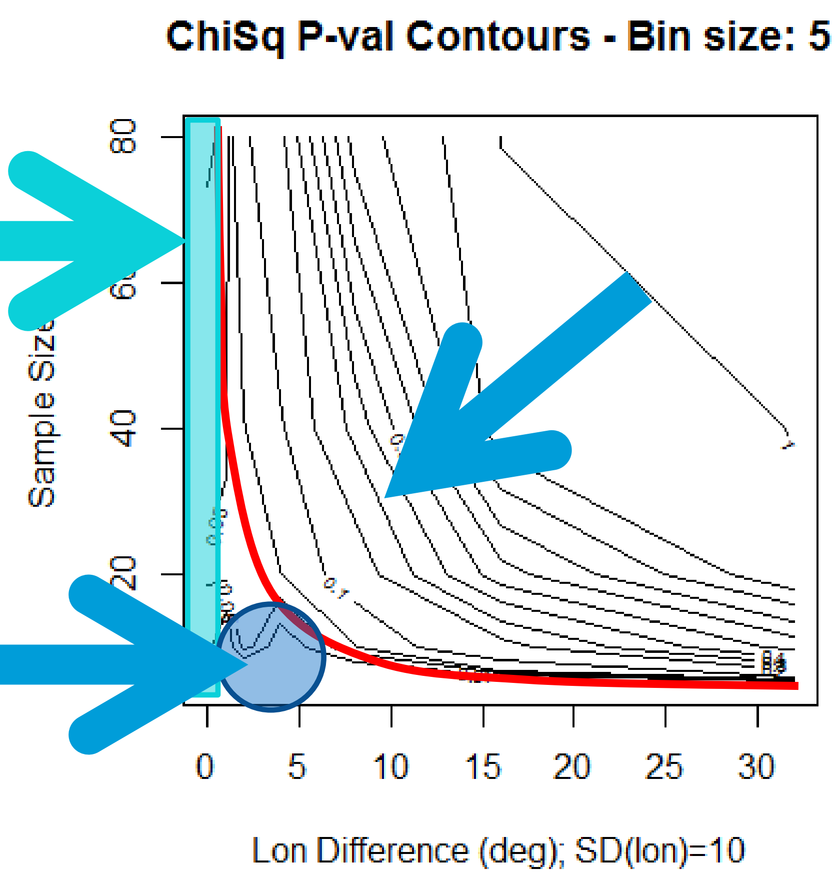
Simulated Indices of Similarity

Type I Errors 5%

- except for <5 tags, then error<5%
- i.e. does not create false positive

Type II Errors (Low Power)

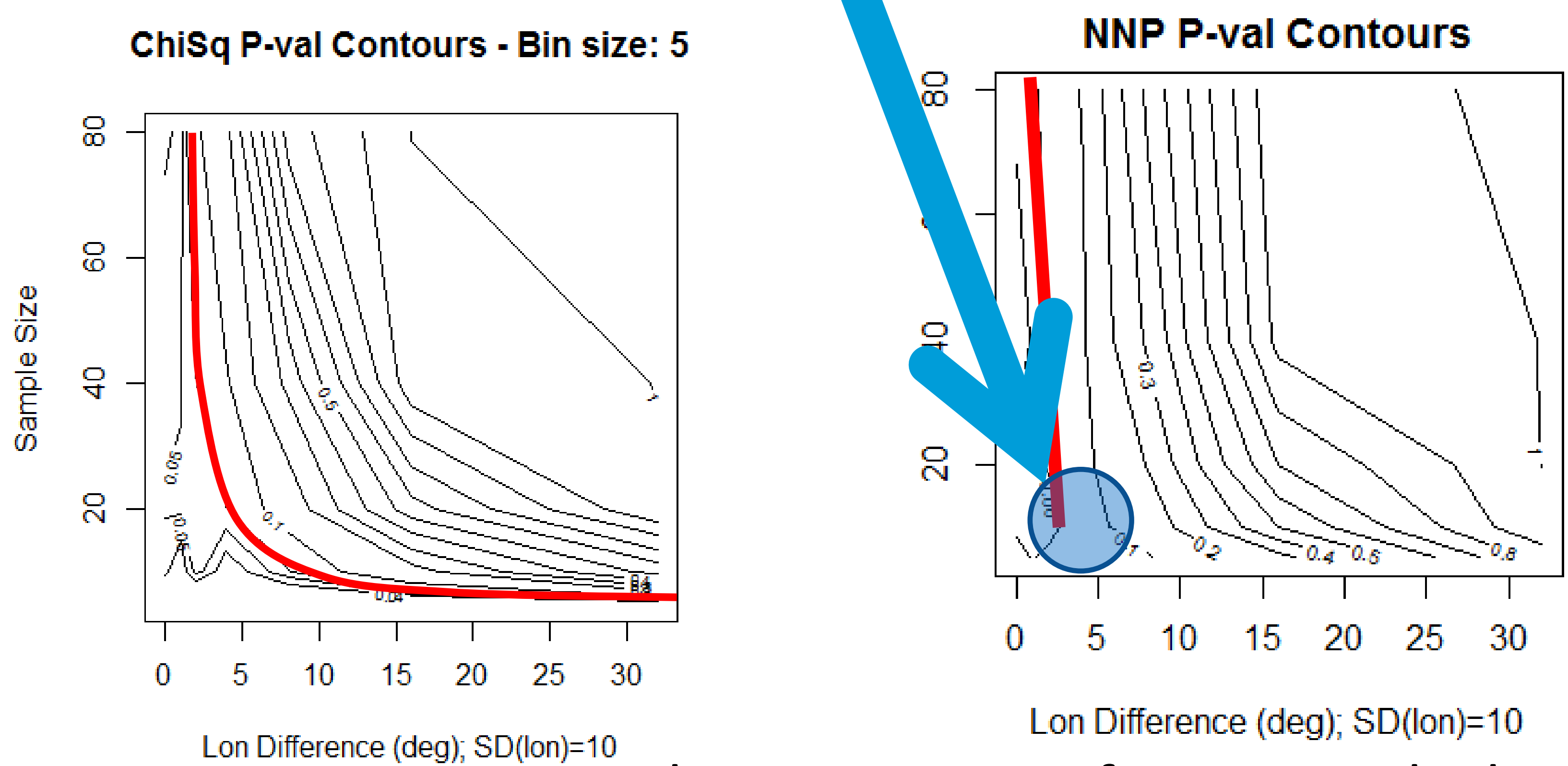
- Cannot identify real differences with few tags, even with many CuSTaRDs



Contours represent the proportion of times in which $p<0.05$ was observed

Comparison of Chi-square and NN-P

Similar, but NN-P more powerful



Contours represent the proportion of times in which p<0.05 was observed (5% expected by chance)

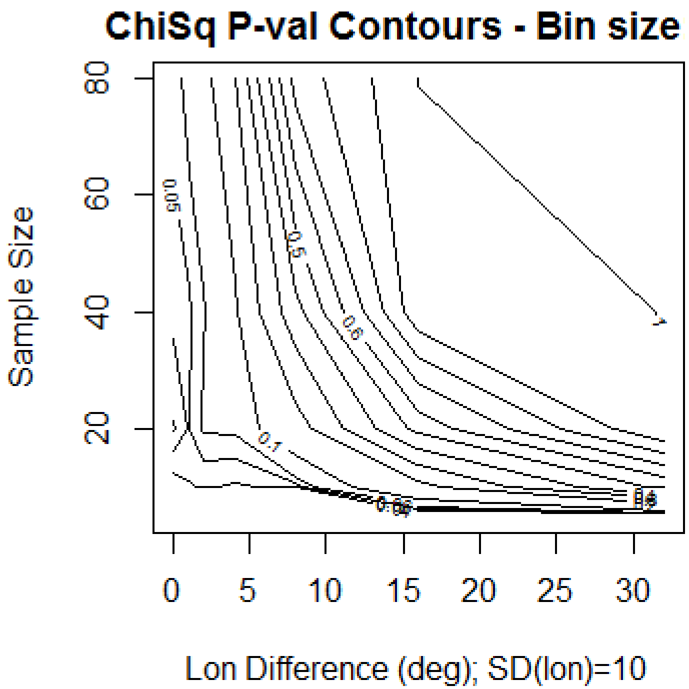
Conclusions from Simulations

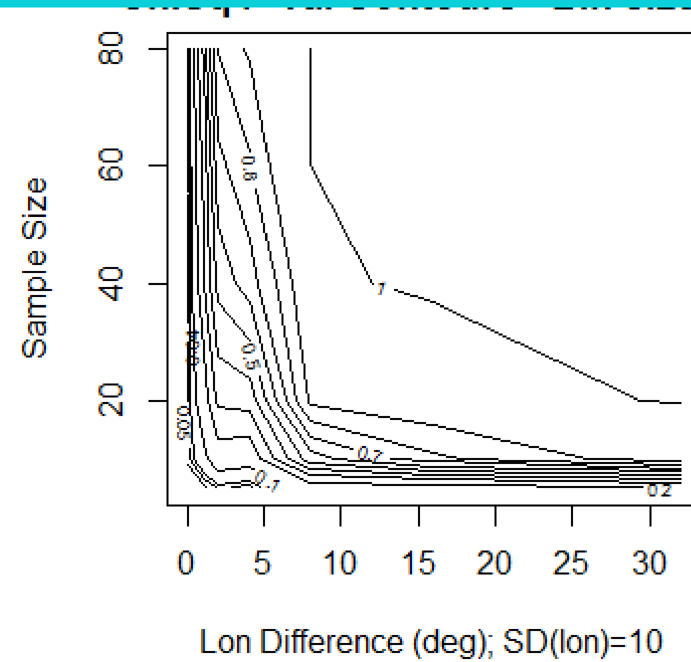
- Chi-square and NN-P produce similar results, generally consistent with visual expectation
- NN-P more powerful
- Sample sizes 5-10 marginal

- Have not found any mechanism to produce false positives at rate >alpha?

P-value Simulations comparing Normal and mixed CuSTARds

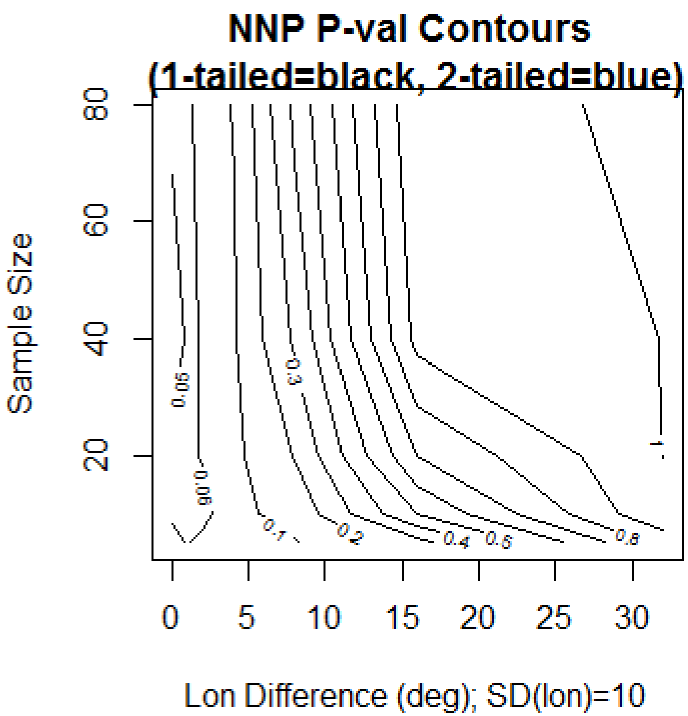
Mixed easier to detect differences, no evidence of false positive

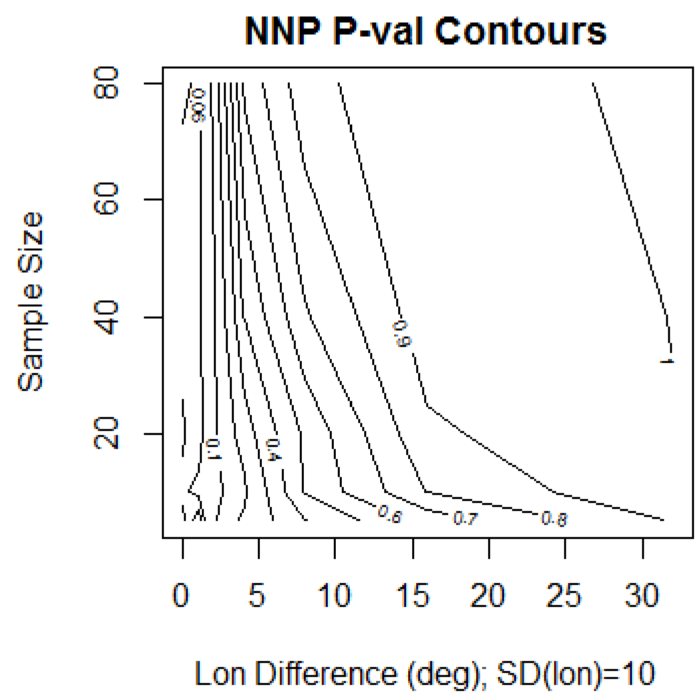




P-value Simulations comparing Normal and mixed CuSTARds

Mixed easier to detect differences, no evidence of false positive

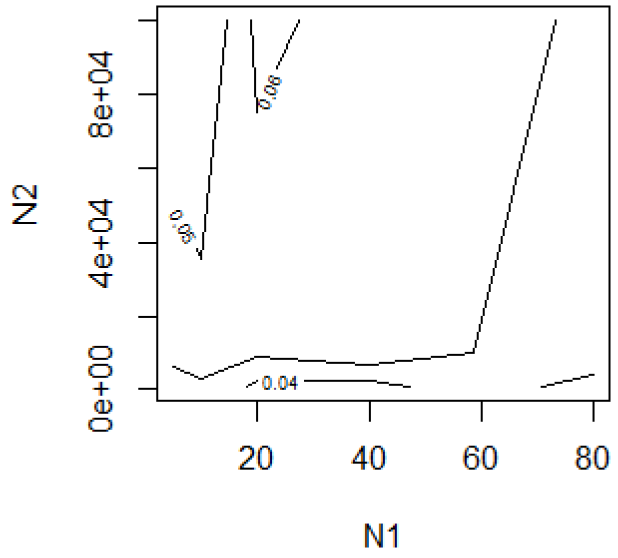




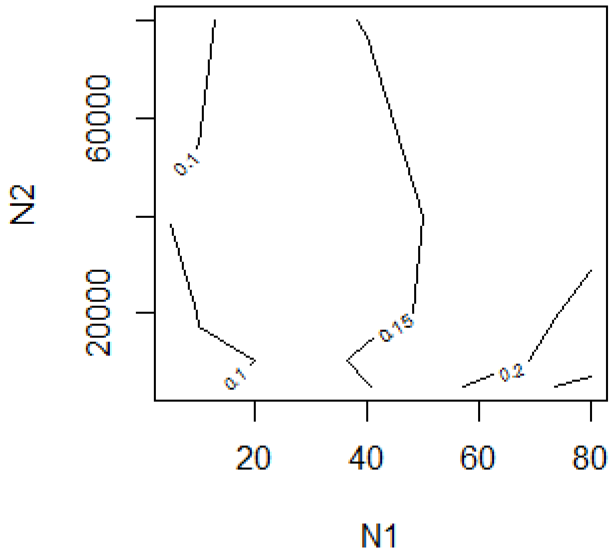
P-value Simulations comparing very big and very small distributions (e.g. Useful for Catch vs tags)

deltaLon=0, no evidence of false positive, no real difference between 1K,10K,100K

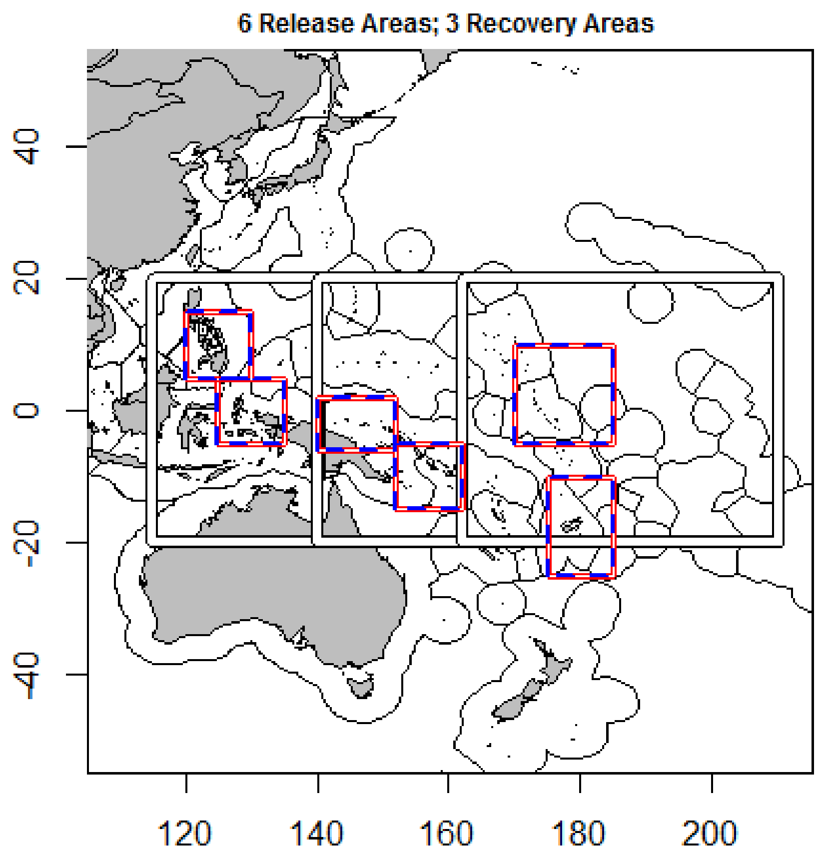
deltaLon=5, maybe there is a difference for N1 >60 with N2?



ChiSq P-val Contours - Bin size: 5

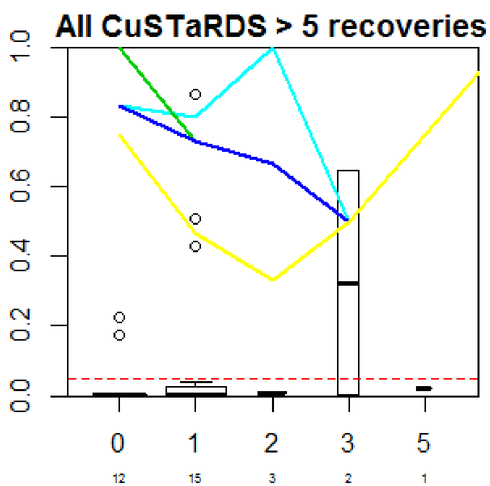


How about three longitudinal areas? (each with two core release locations)



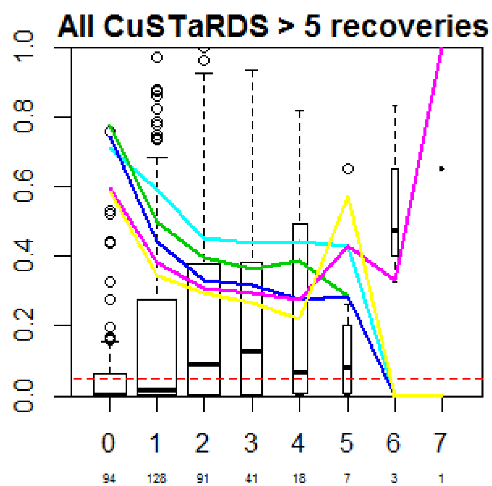
How about three longitudinal areas? (each with two core release locations)

Western



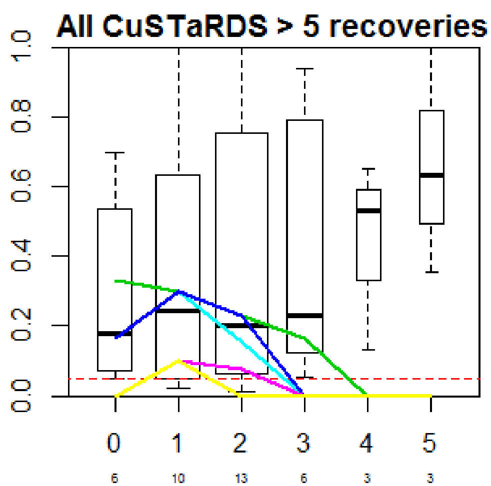
Poor Mixing

Central



Poor Mixing

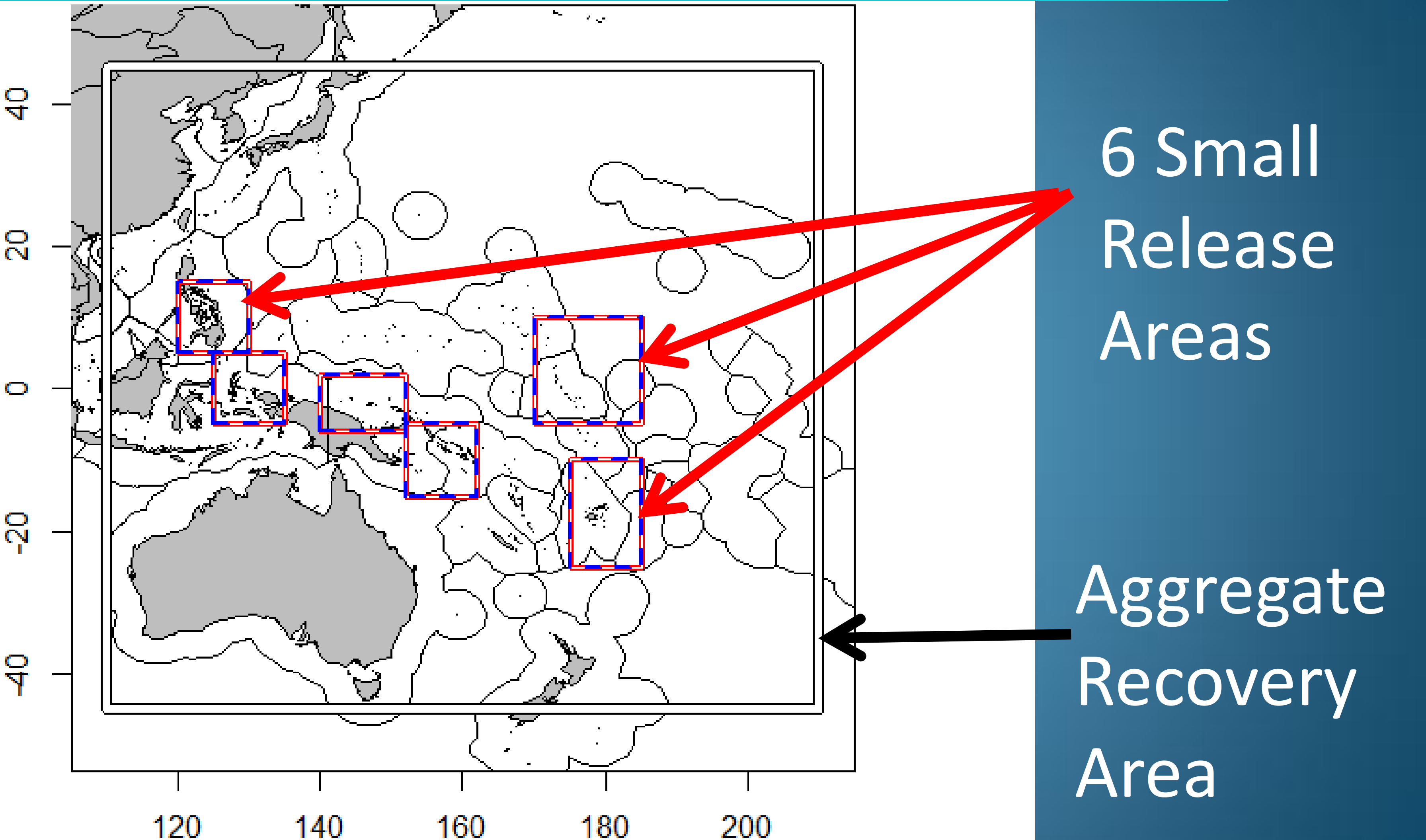
Eastern



Higher mixing in pelagic waters?

CuSTaRD Example 1:
Do we really need spatial Structure in the WCPO SKJ Stock Assessments?

Evaluating mixing across the whole WCPO



816 Paired Recovery Distributions

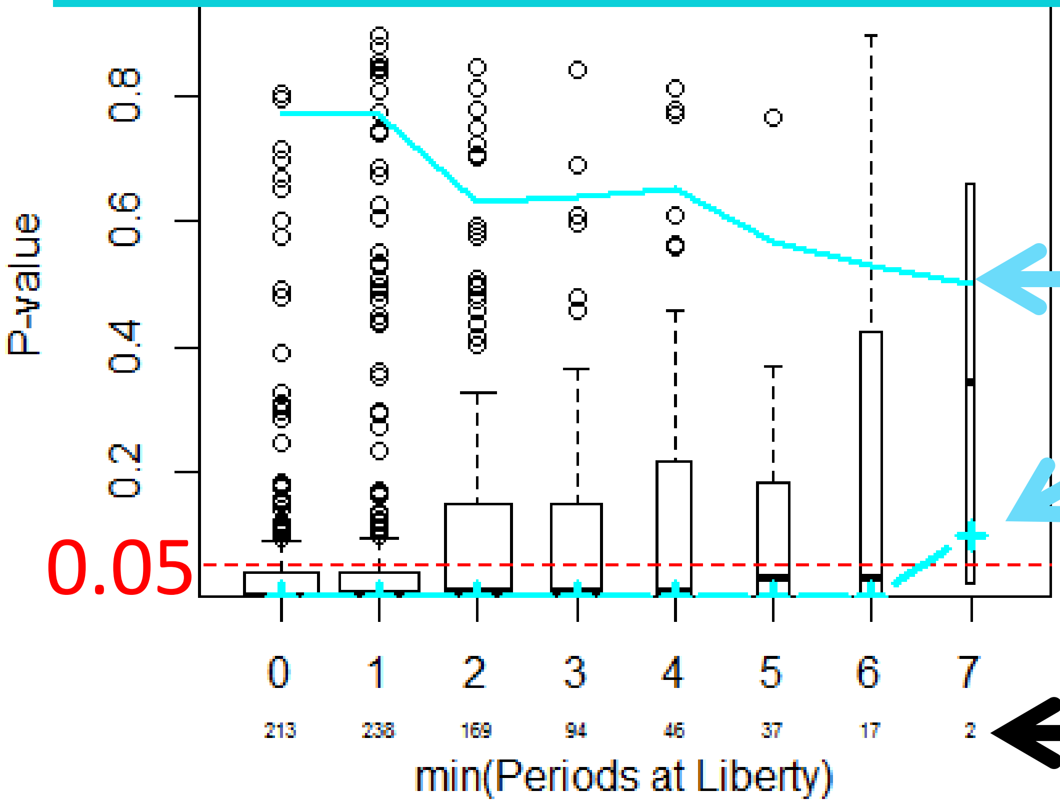
- By Release Area:
  - 174 CUSTARDs from the same release area
  - 642 CUSTARDs from different release areas

- By Size class:
  - 1 smallest size
  - 197 middle size
  - 618 largest size

- By Tagging Programme
  - 2000s = 357
  - 1990s = 419
  - 1970s = 40

Actual Results for Aggregate Case (816 CuSTaRDs)

Strong evidence of incomplete-mixing for at least 6 quarters



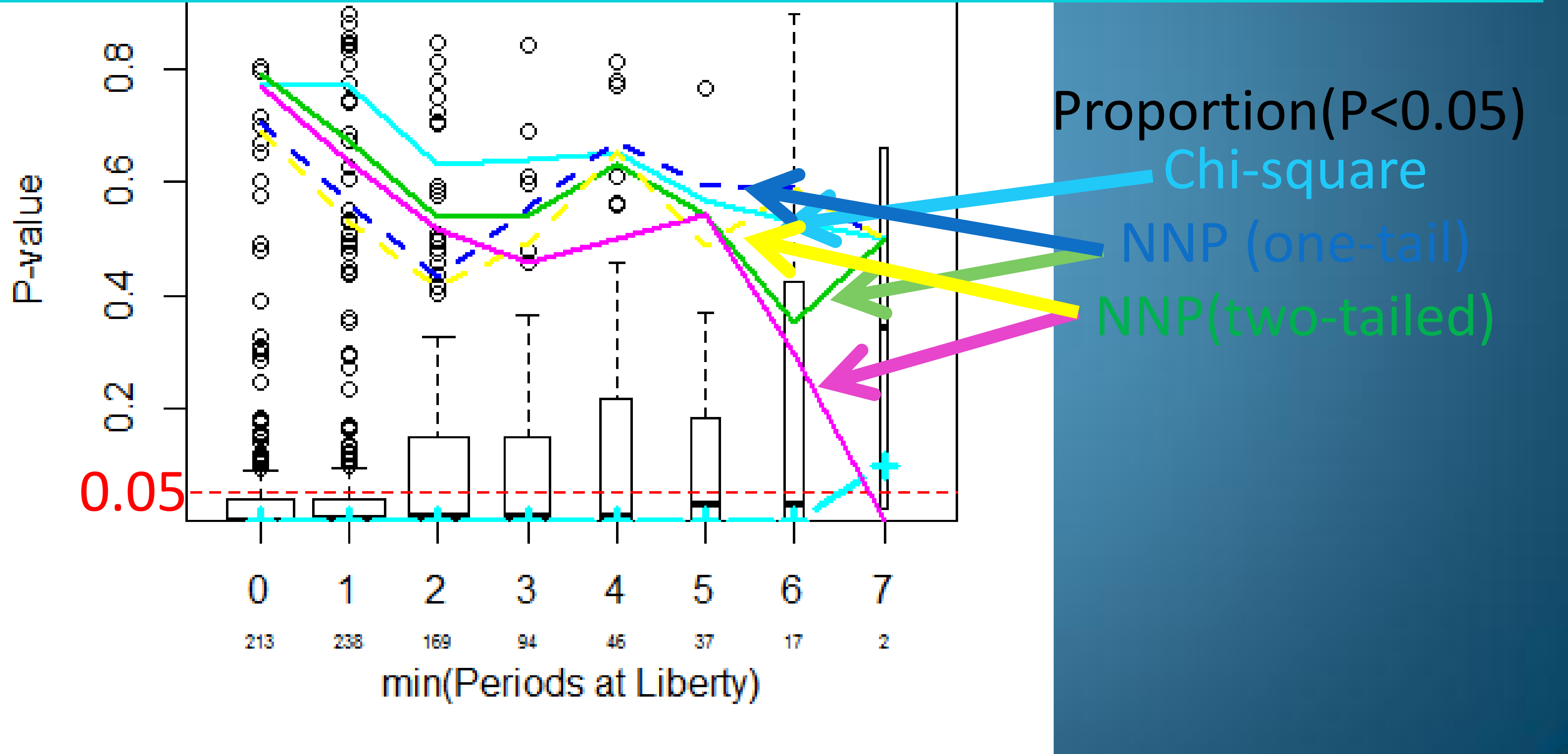
Proportion(P<0.05)

Binomial probability of obtaining this proportion by chance

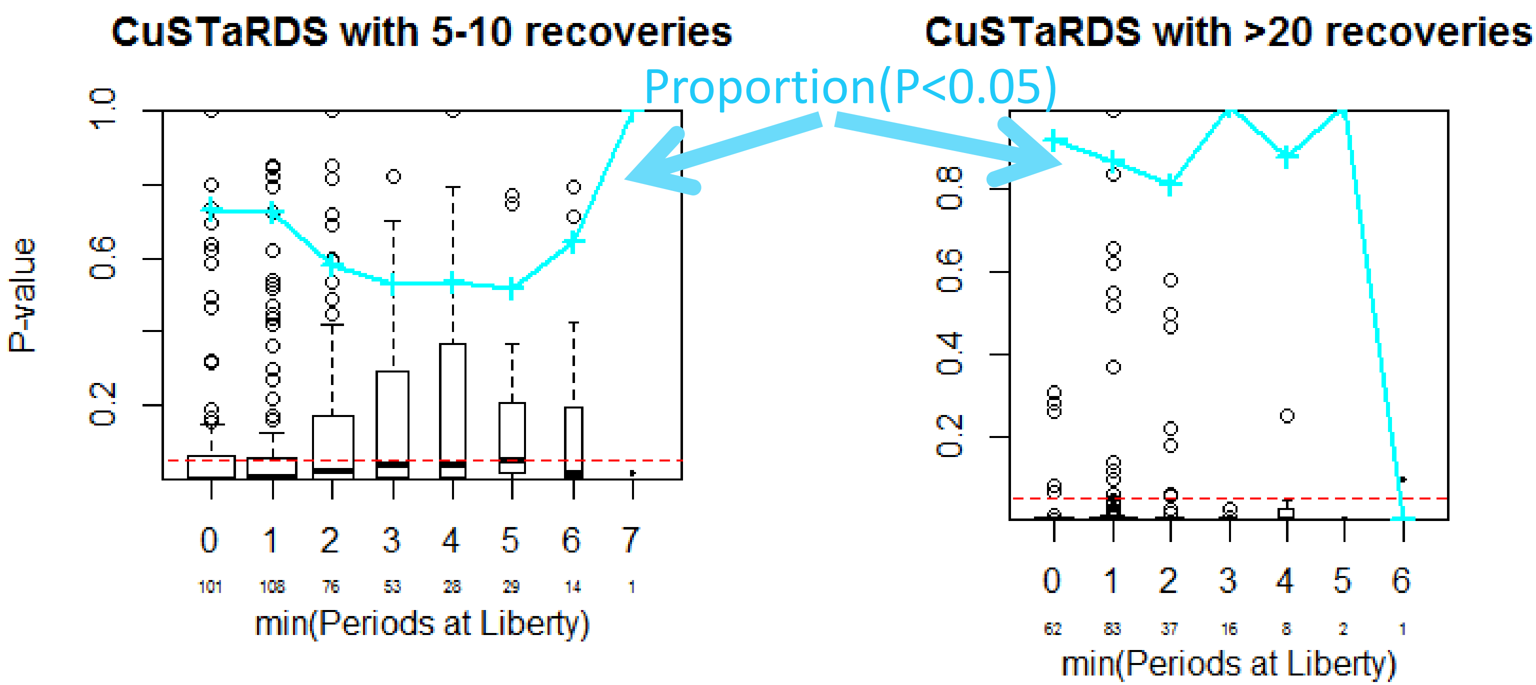
N (CuSTaRDS)

Chi-square and NNP Comparison

Different Methods of comparing distributions often yield different results for an individual CuSTaRD, but the overall result is consistent

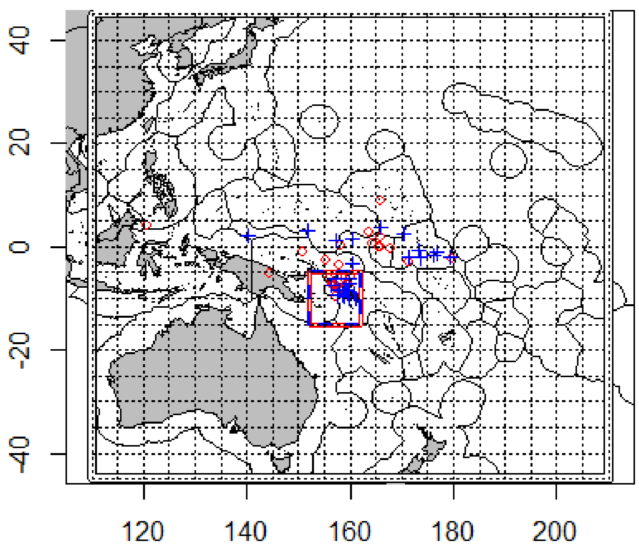


Small tag recovery sample sizes probably reduce the power to detect differences

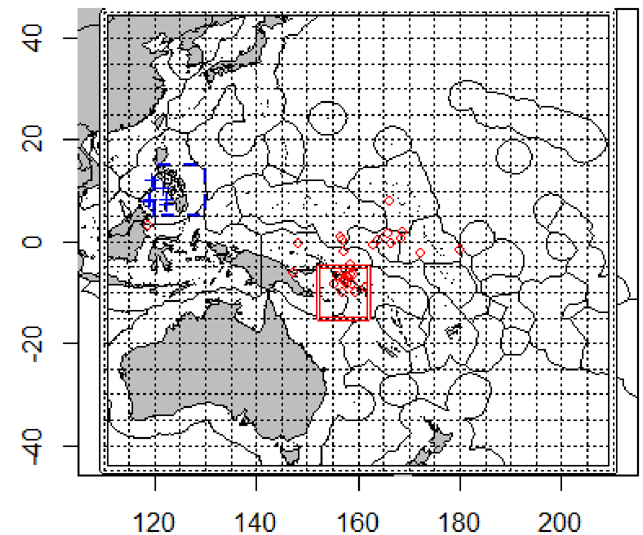


Recoveries from the same release area (separated in time) are probably mixed more than releases from different release areas

Same release area



Different release area



Conclusions for Aggregate WCPO CUSTARD Analysis

- Despite various factors that reduce power, even 6+ quarters after release, there is compelling evidence that SKJ are not well-mixed across the WCPO
- Evidence is stronger with multiple release locations, larger recovery numbers and higher number of CuSTaRDs
- Hard to argue that a spatially-aggregated WCPO SKJ assessment would conform to the standard tag mixing assumptions

So how big are the F-biases really?

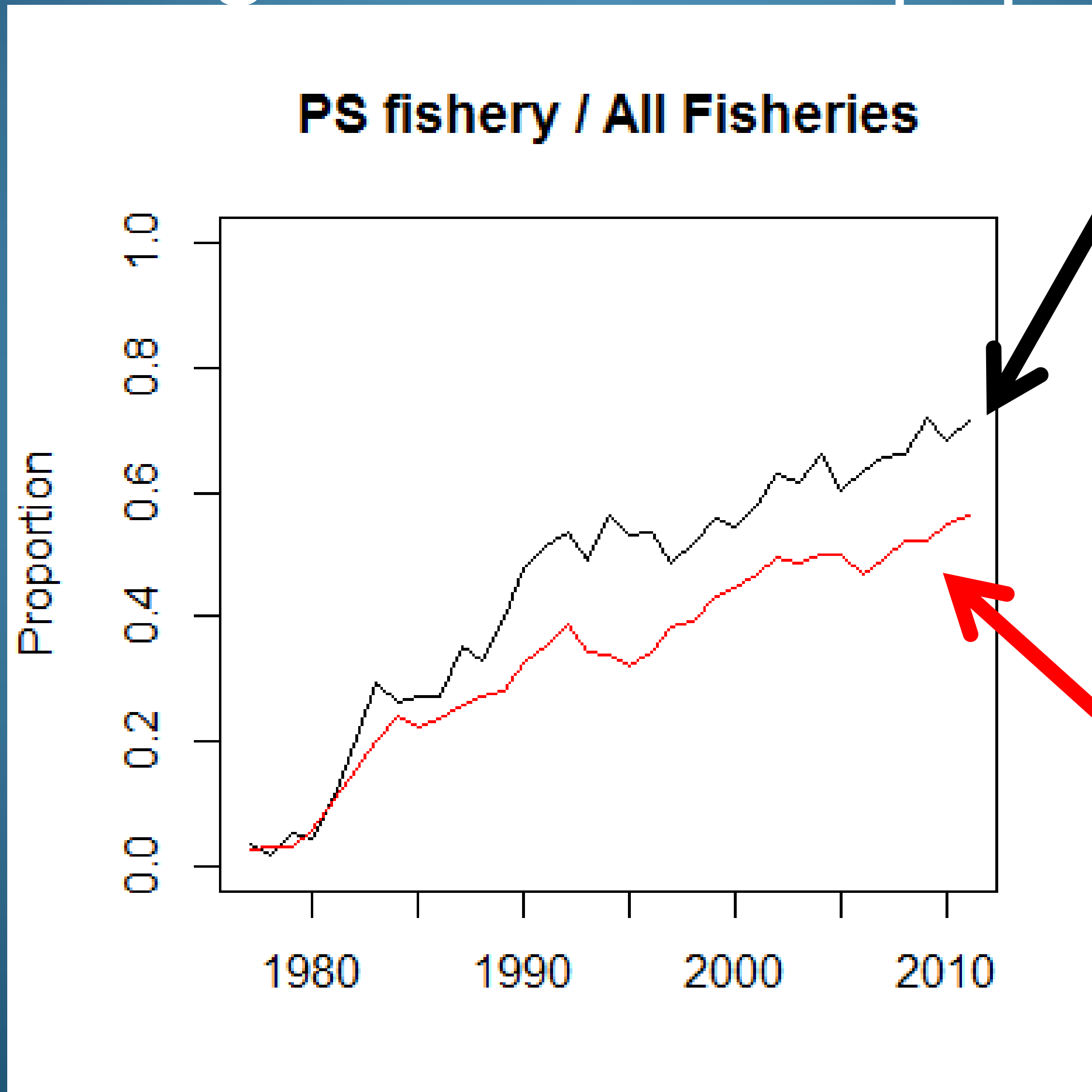
- Cannot answer this without strong assumptions about the distribution of catch (or effort) in relation to the fish population.

- Mechanistic simulation testing of alternative plausible spatial distributions of the population?
  - e.g. SEAPODYM?

Could the PS fleet be fishing in proportion to population density?

- Incentive for high CPUE
- Other constraints due to landing ports, access rights, management measures
- Observable differences among fleets
- Can we quantify biases in F?
  - Can compare F estimates by:
    - Subsetting release events
    - Subsetting fisheries?
- But probably need credible mechanistic simulation framework to get at F biases

How likely that the (tag-seeded) PS fishery is covering the whole SKJ population?



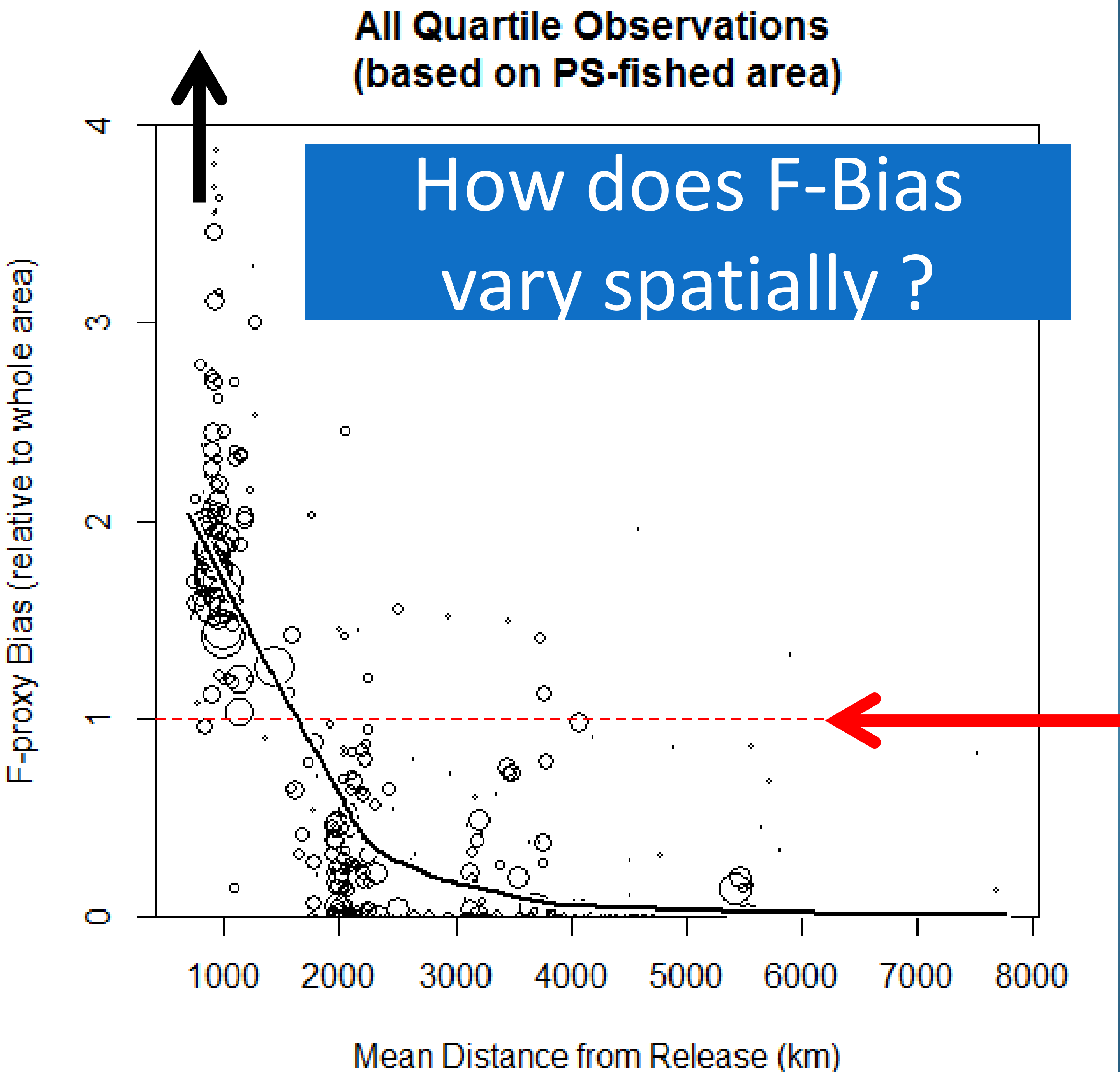
Total catch

Number of strata with positive catch

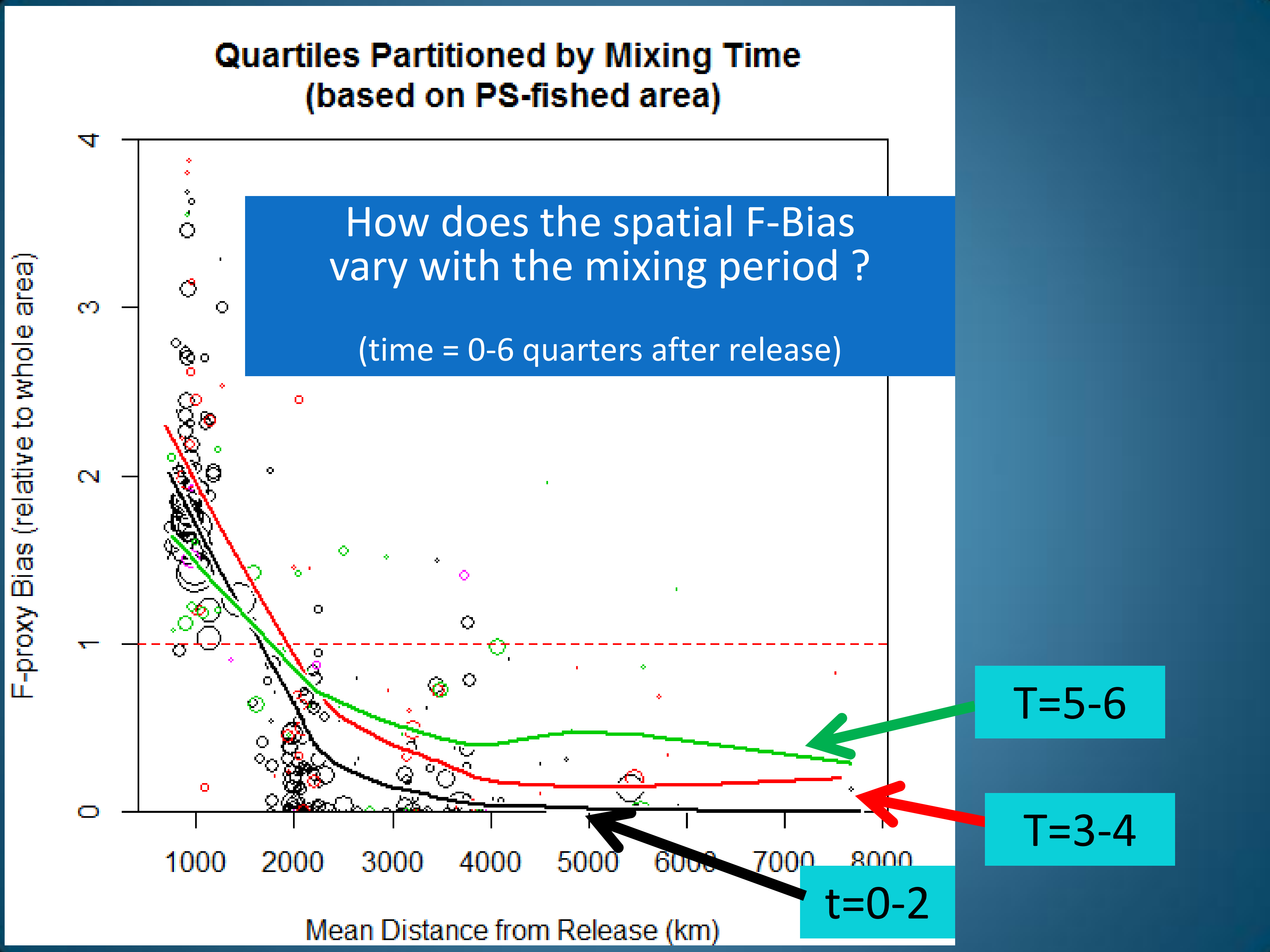
What happens with real data?

- Pacific Skipjack tuna tagging history
- PTTP 2006-2012(+)
- RTTP 1989-92
- SSAP 1977-82
  - Include small-scale programmes administered through SPC
- JPN 1970s +

How does F-Bias vary spatially ?

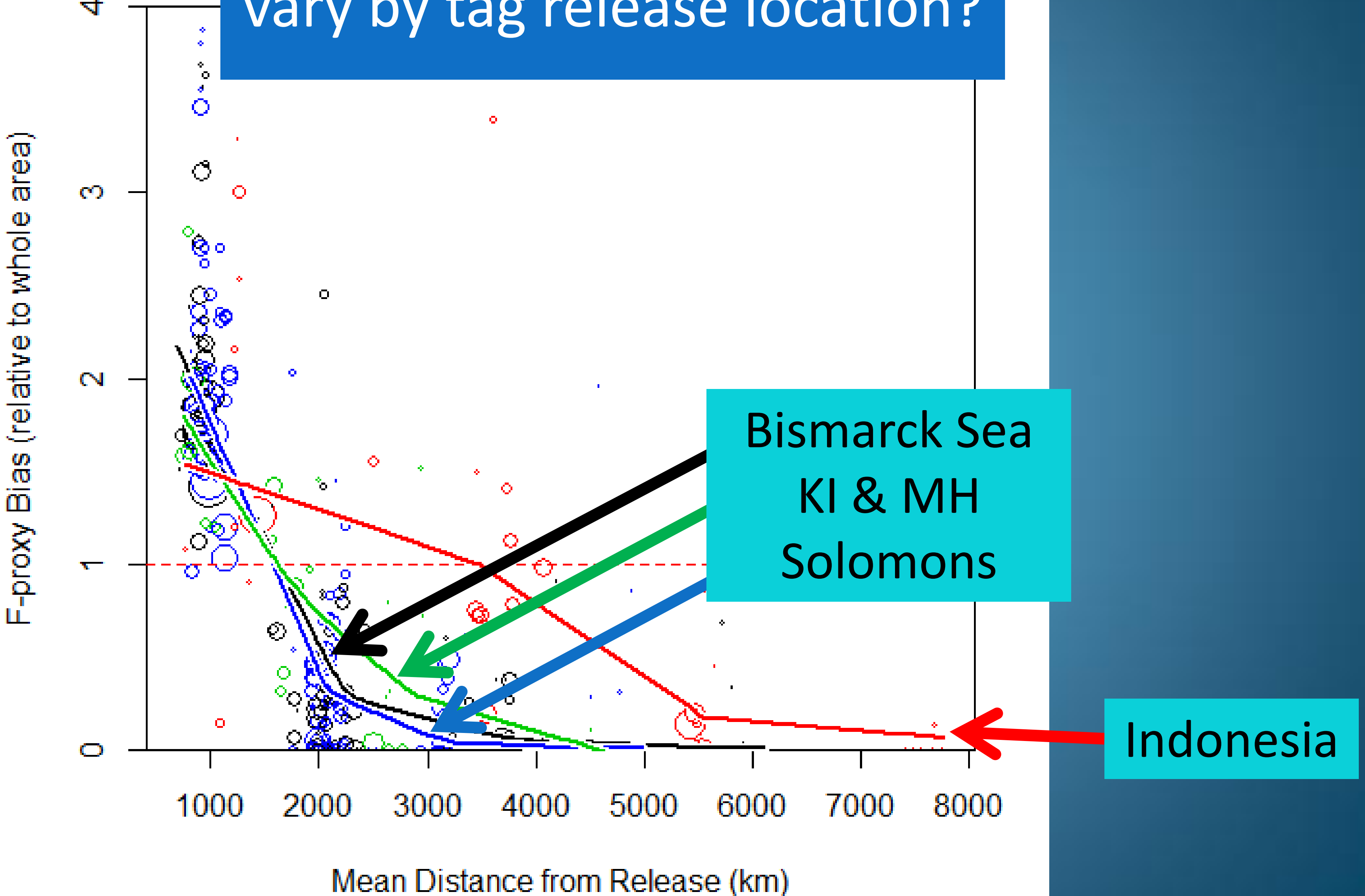


F-proxy for the whole region

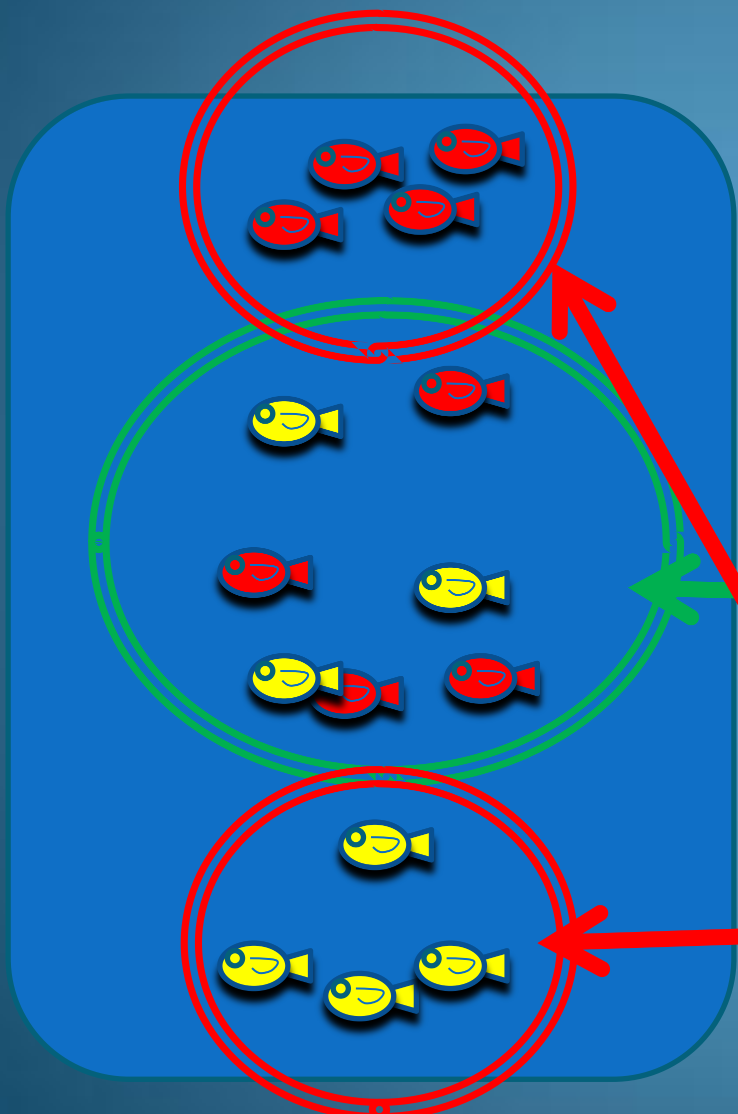


Quartiles Partitioned by Mixing Time
(based on PS-fished area)
How does the spatial F-Bias vary with the mixing period ?
(time = 0-6 quarters after release)
F-proxy Bias (relative to whole area)
0
1
2
3
4
1000
2000
3000
4000
5000
6000
7000
8000
Mean Distance from Release (km)
T=5-6
T=3-4
t=0-2

How does the spatial F-Bias vary by tag release location?



Recovery distributions might appear to be mixed when the tags are not mixed,e.g.



- Release Event 1
- Release Event 2

1) Tags are only reported from a restricted area

2) Tags caught but not reported