Exploratory analysis of yellowfin tuna longine catch and effort data using VAST

IOTC 15th Working Party on Methods (WPM)

Seychelles 24-26 October 2024

Information Paper 01

Adam Langley (IOTC Consultant)

Introduction

- Review of IO YFT stock assessment in 2023, presented to WPTT25.
- Key recommendations included:

"Alternative abundance indices required to corroborate or replace LL CPUE. Review assumptions associated with LL CPUE indices. Global tuna RFMO meeting to progress common issues with LL CPUE".

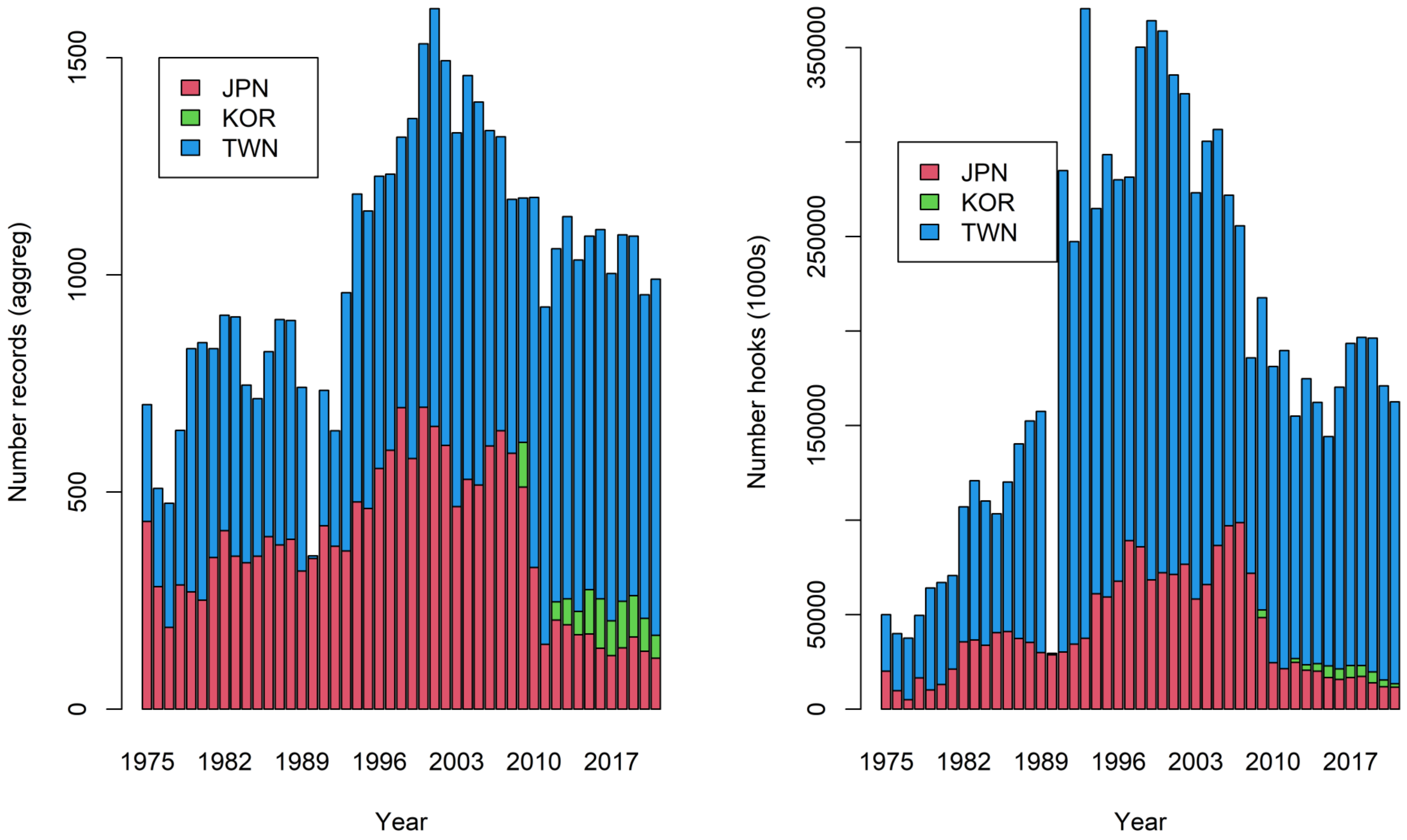
- Currently, LL CPUE indices derived from delta-lognormal GLM of DWFN logsheet data, including species catch clustering. Investigate application of VAST (Vector Autoregressive Spatio-Temporal Model) for CPUE modelling.

CPUE analysis – Methodology

- Exploratory spatial temporal analysis of LL CPUE data using VAST (version 3.11.1).
- LL 5*5 aggregated catch (YFT_NO) and effort (num_hooks).
- Limited to three DW fleets: JPN, KOR, TWN.
- Recent period 2000-2021.
- Entire Indian Ocean domain.
- Separate VAST models for each season (4). Occurrence (binomial) and magnitude (lognormal *not discrete*). Simultaneously estimate spatial variation and spatial-temporal variation.
- Extract overall index and regional indices (4).
- Model includes FLEET as a *catchability* explanatory variable for the lognormal model. Temporal variation in spatial effect (random effects).
- *Pred ~ YR + LAT_LONG + as.factor(FLEET) + HOOKS +* LAT_LONG_t
- Comparable region specific GLMs using the equivalent data set (direct comparison).

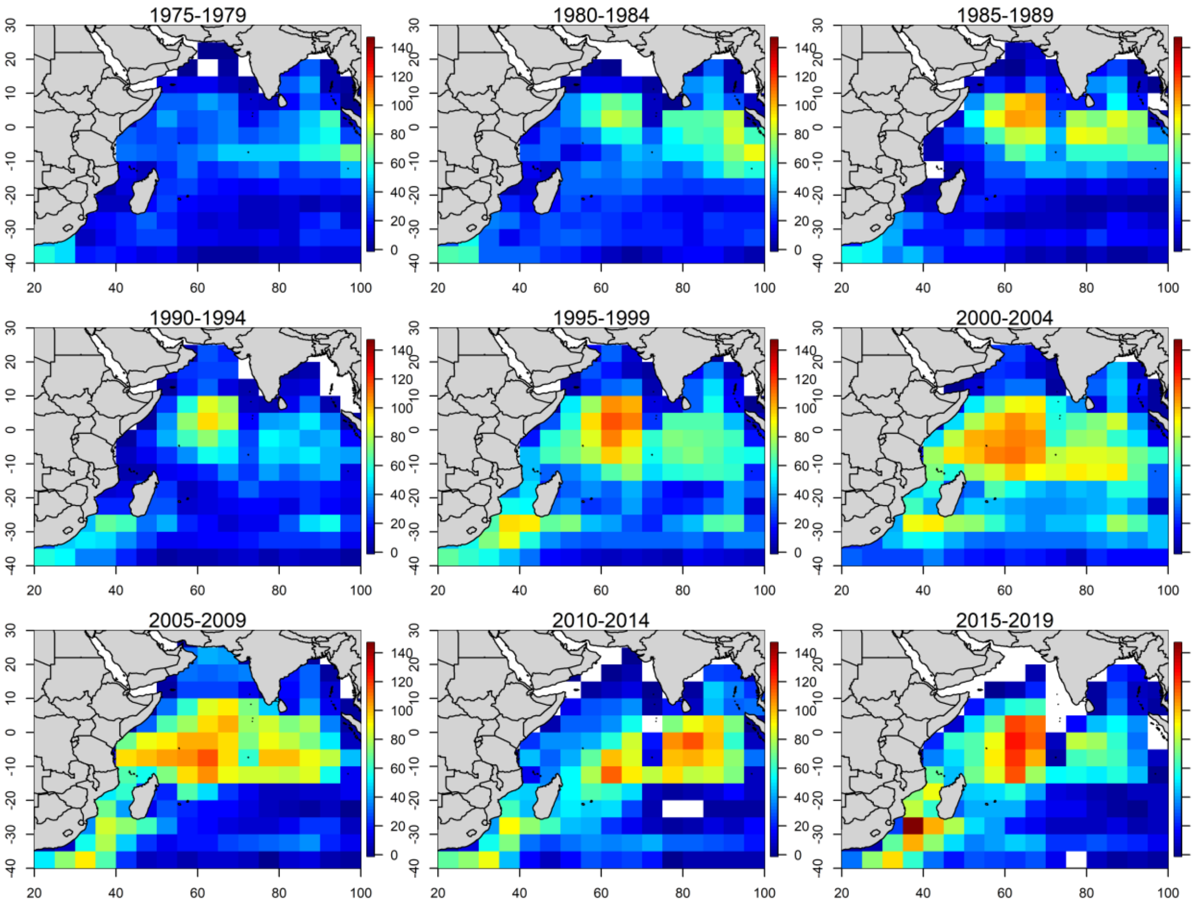
Pred ~ as.factor(YRQTR) + as.factor(LAT_LONG) + as.factor(FLEET) + HOOKS

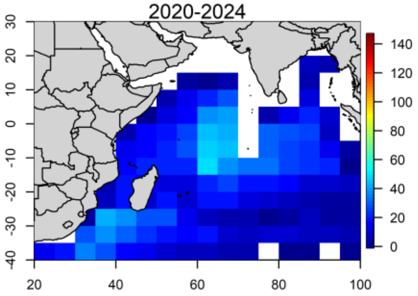
INPUT DATA


JPN
KOR
TWN
Number records (aggreg)
0
500
1000
1500
1975
1982
1989
1996
2003
2010
2017
Year
JPN
KOR
TWN
Number hooks (1000s)
0
50000
150000
250000
350000
1975
1982
1989
1996
2003
2010
2017
Year

INPUT DATA

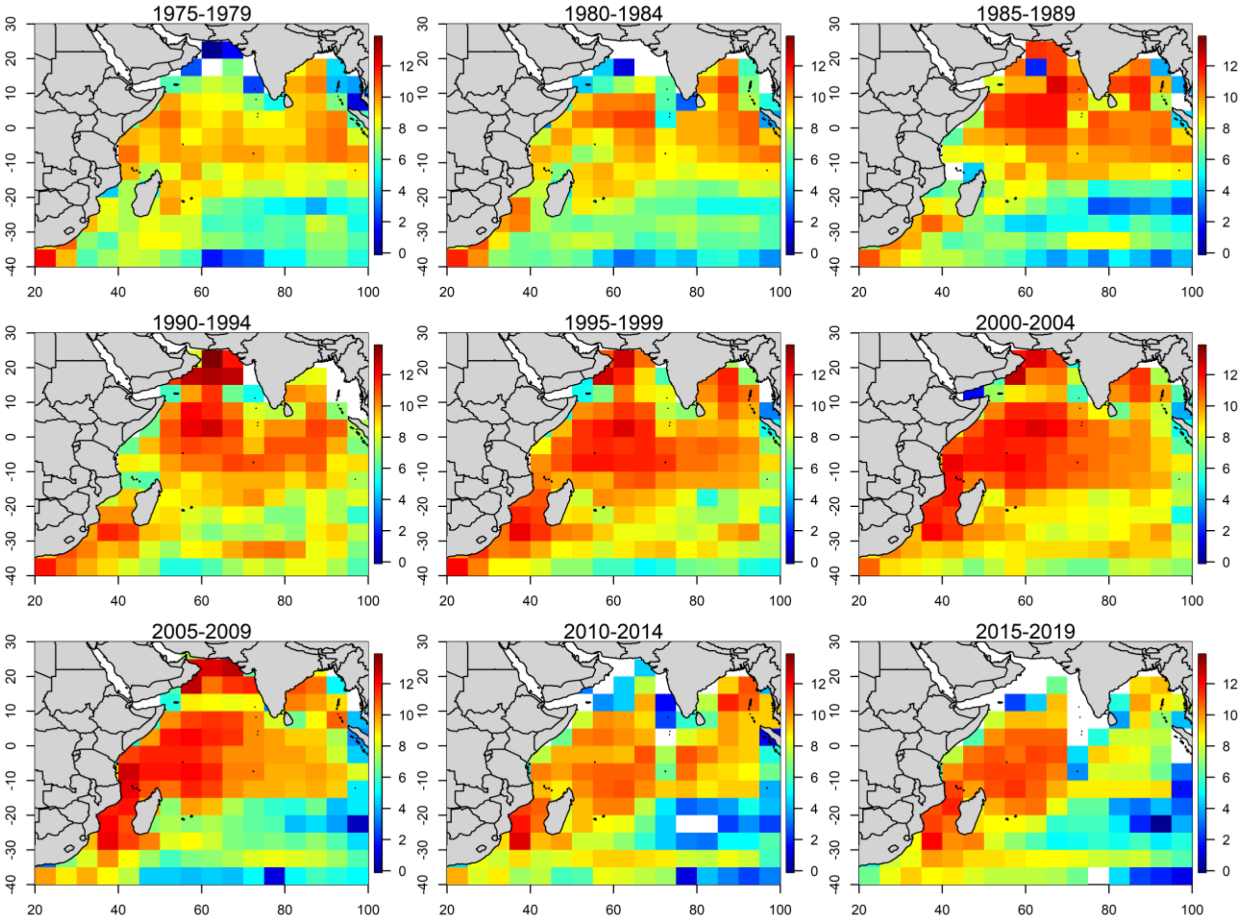
Records

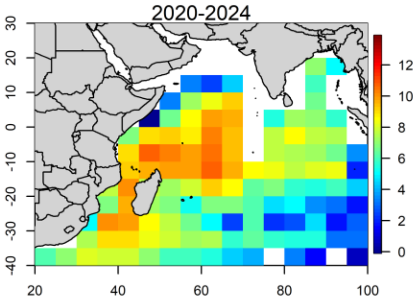




Limited records in North East IO in recent years.
Closure of Maldives LL fishery in 2019.

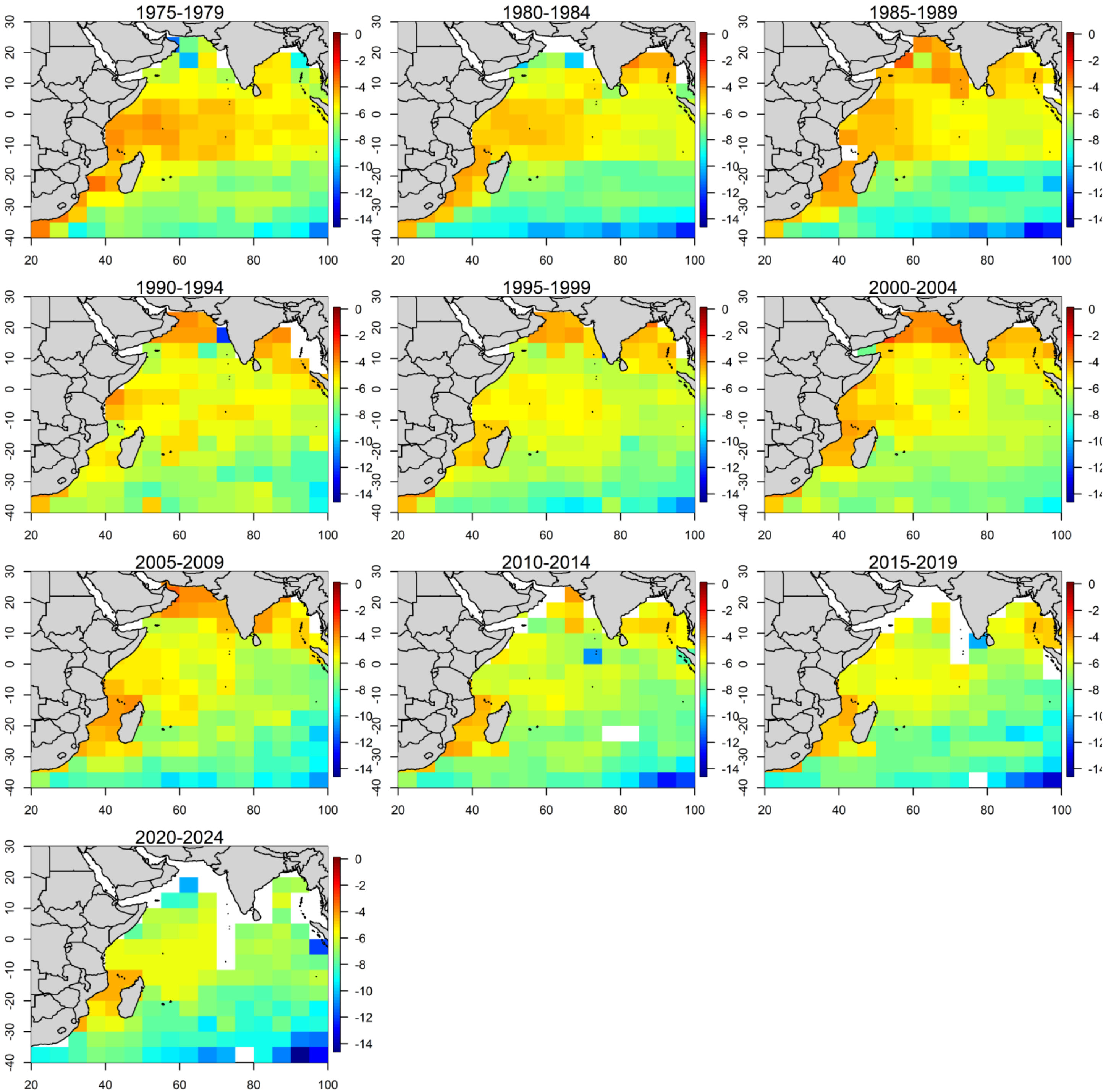
YFT Catch (log Number)





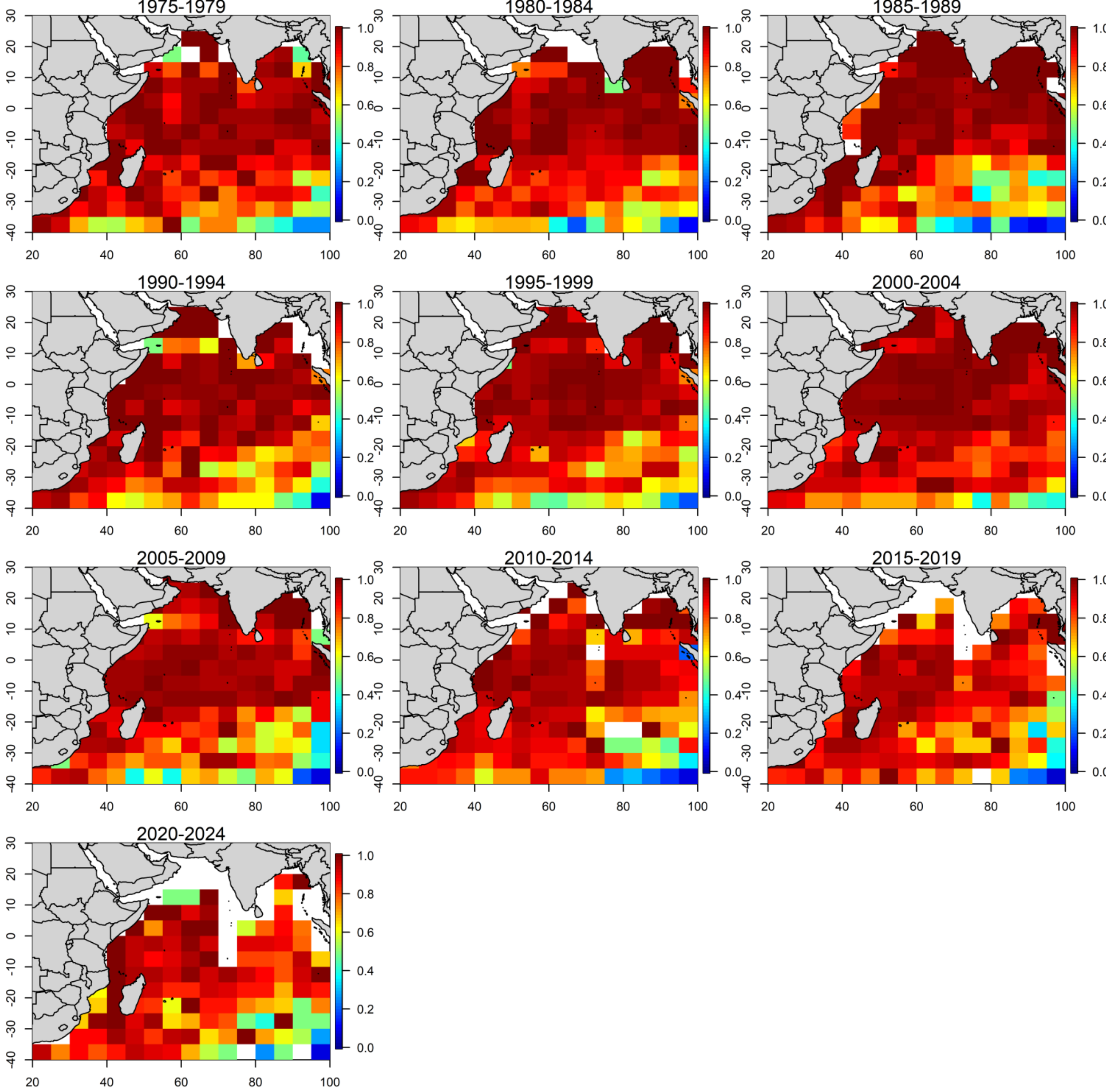
Nominal YFT CPUE

Log Num/Hooks

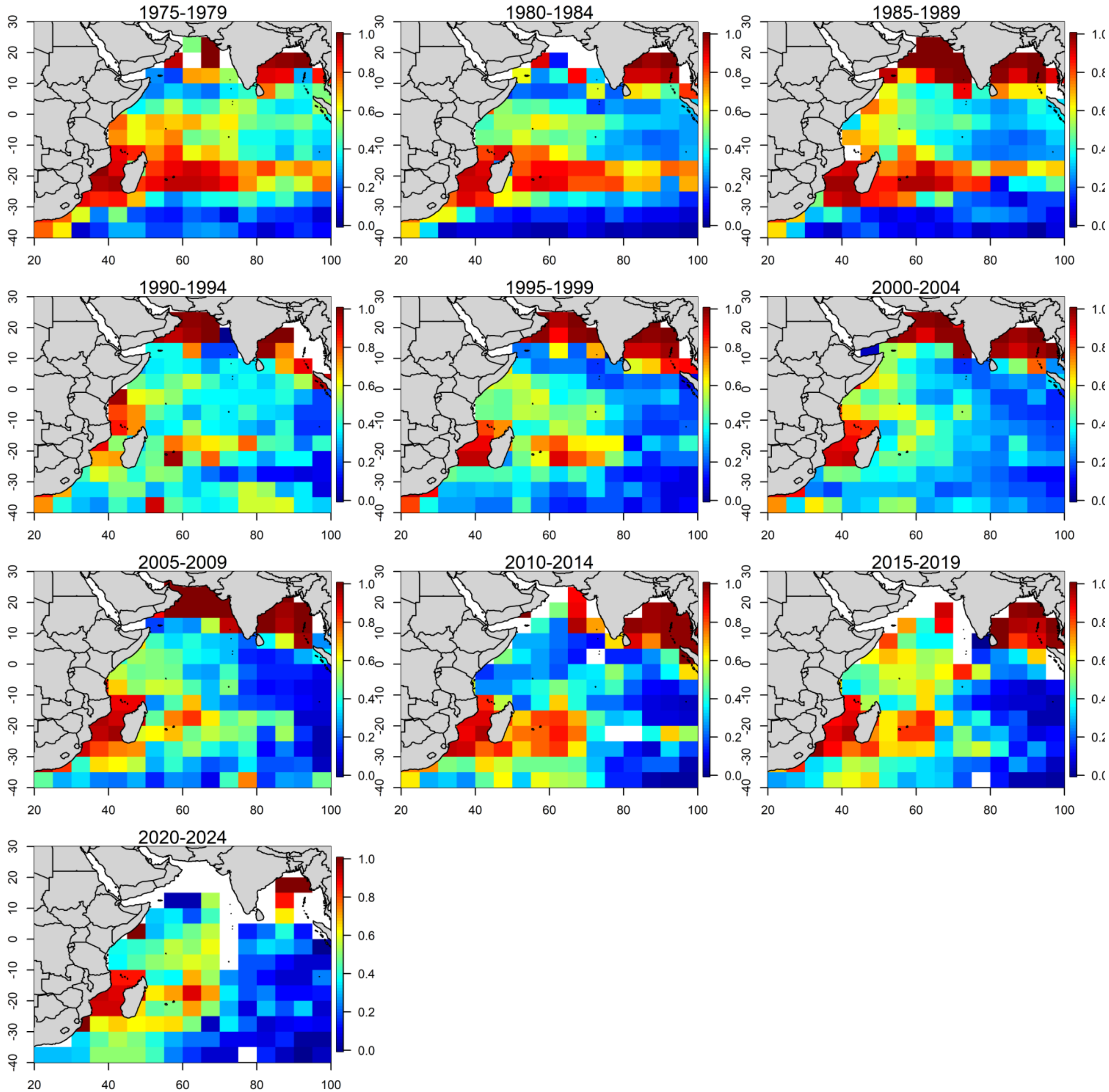


Frequency of YFT catch (per record)

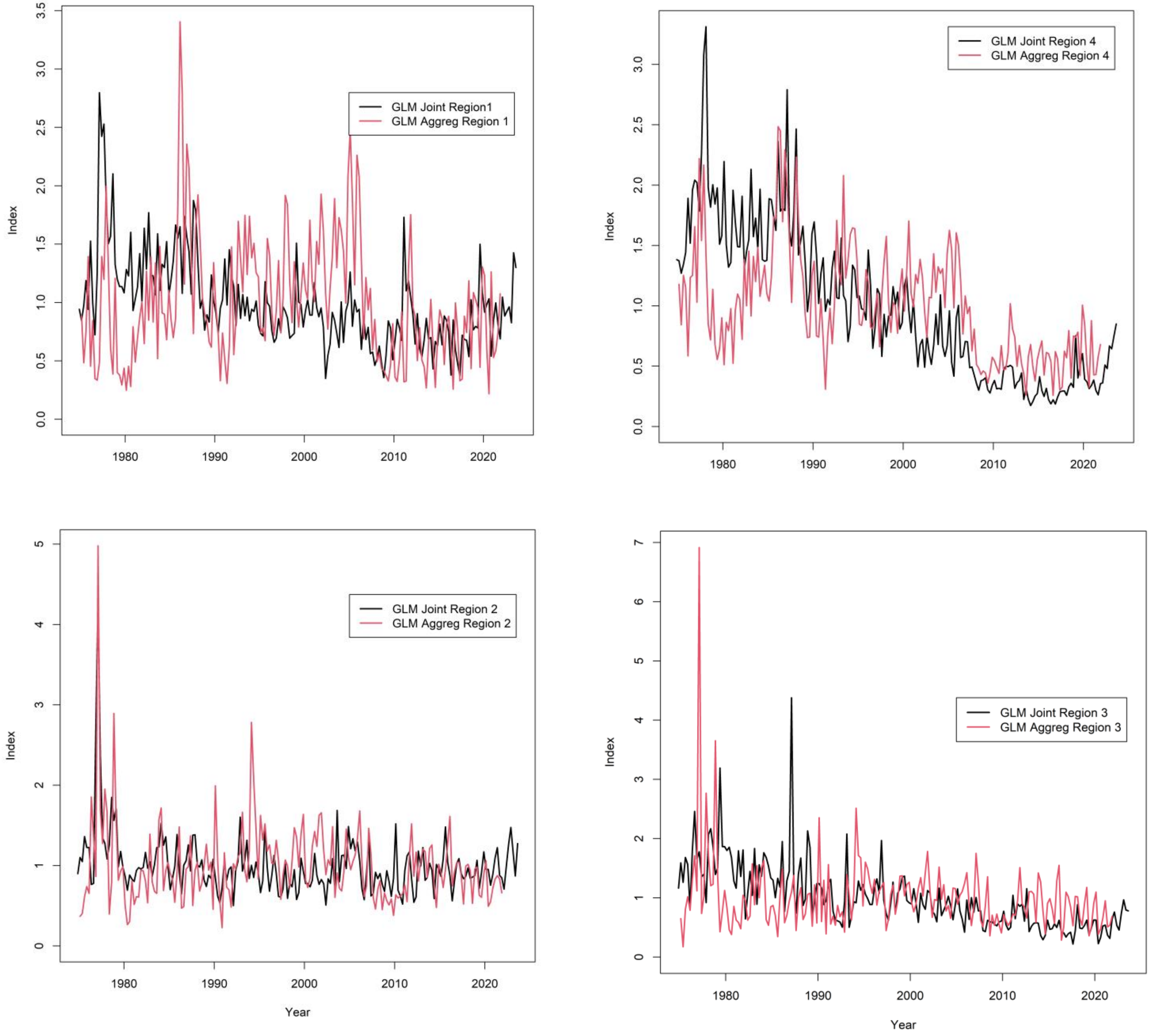
Proportion of records with greater than zero YFT.



Proportion of YFT in YFT+BET catch (per record)



GLM aggreg 5*5 vs GLM Joint Logsheet (2024)



GLM indices derived from aggregated data differ from accepted GLM indices (Joint).

Main difference is during 1975-1985 (all regions) and for Region 1.

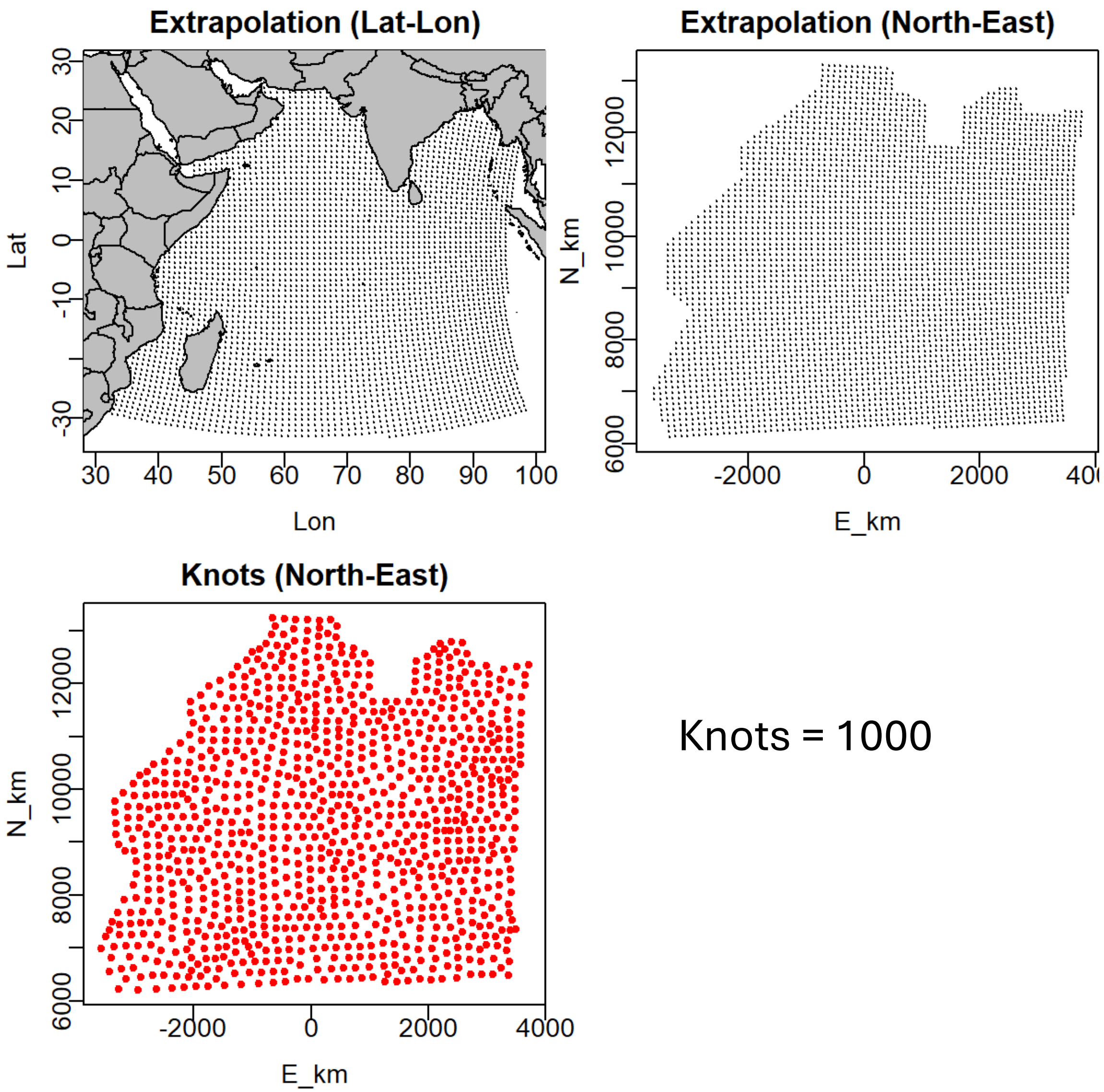
GLM aggreg 5*5 not presented as alternative abundance indices. Instead used for contrasting results with VAST.

VAST model

- *Comp1 ~ YR + LAT_LONG + as.factor(FLEET) HOOKS*
- *Comp2 ~ YR + LAT_LONG* + LAT_LONG_t + *as.factor(FLEET) + HOOKS*
- *YR* fixed effects, LAT_LONG_t random effects,
- FLEET is a catchability effect.

$number_of_coefficients

Total	Fixed	Random
26694	54	26640



Knots = 1000

More precisely, the spatial variation terms are modeled as Gaussian Markov random fields with correlations over 2 spatial dimensions and among density and condition (Thorson & Barnett 2017):

$$\text{vec}(\mathbf{\Omega}) \sim GRF(0, \mathbf{R} \otimes \mathbf{V}_{\omega}) \qquad (7)$$

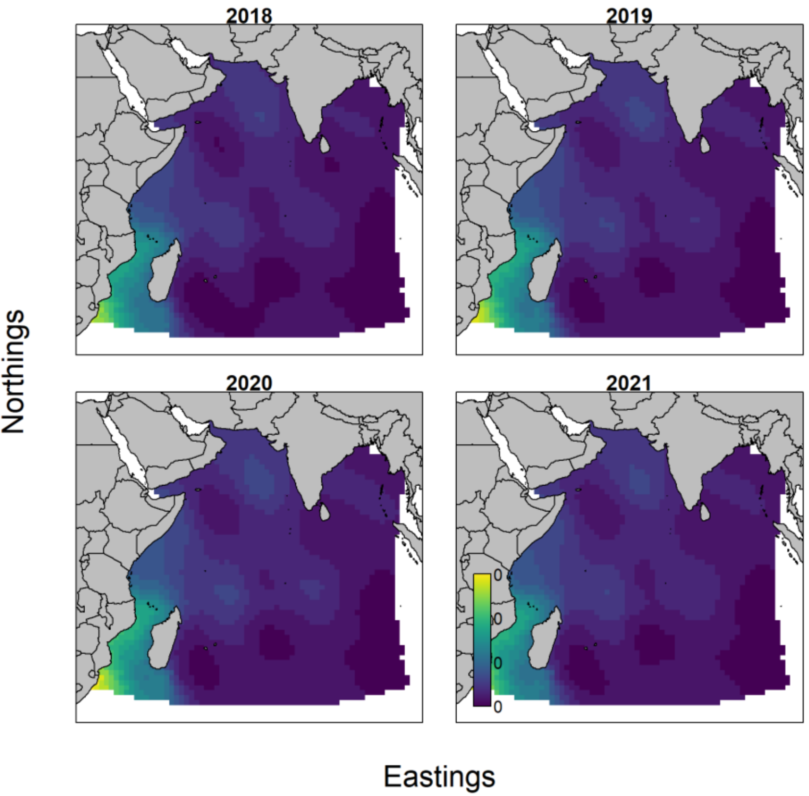
where $\mathbf{\Omega}$ is a matrix composed of $\omega_f(s)$ at every location and category of variables (density and condition); $\mathbf{R}$ is a matrix of correlations among locations s, which is calculated from a Matérn function given an estimated decorrelation rate and a transformation matrix representing geometric anisotropy (Thorson et al. 2015); $\otimes$ is the Kronecker product; and $\mathbf{V}_{\omega}$ is the correlation among density and condition:

$$\mathbf{V}_{\omega} = \mathbf{L}_{\omega}\mathbf{L}_{\omega}^{T} \qquad (8)$$

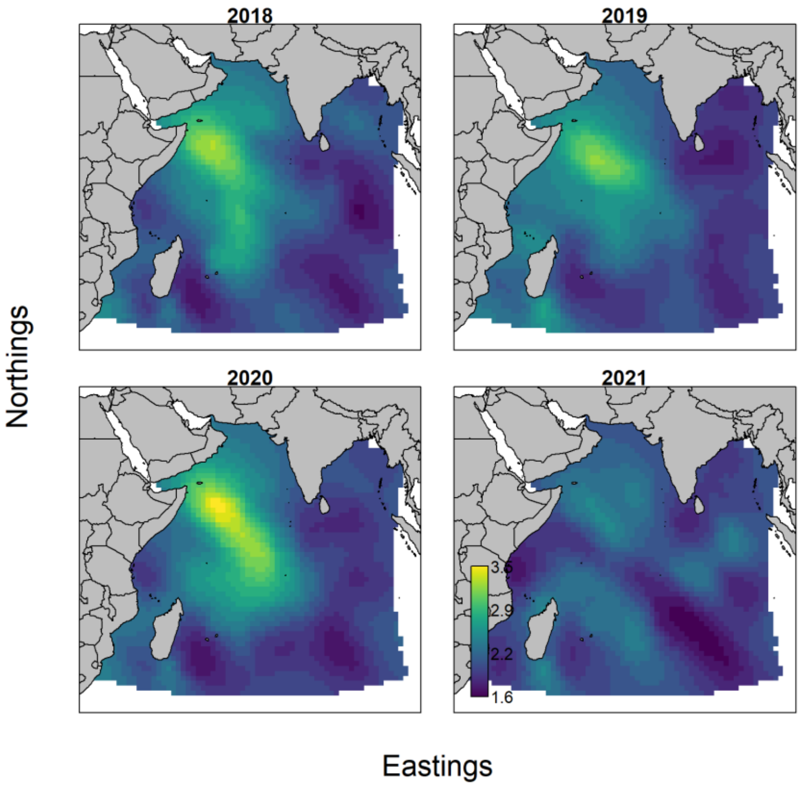
where $\mathbf{L}_{\omega}^{T}$ is the transpose of $\mathbf{L}_{\omega}$. Spatio-temporal variation terms are fit independently to each year and are modeled as Gaussian Markov random fields with Matérn covariance, similarly to spatial variation terms (Thorson & Barnett 2017).

VAST model predictions (example)

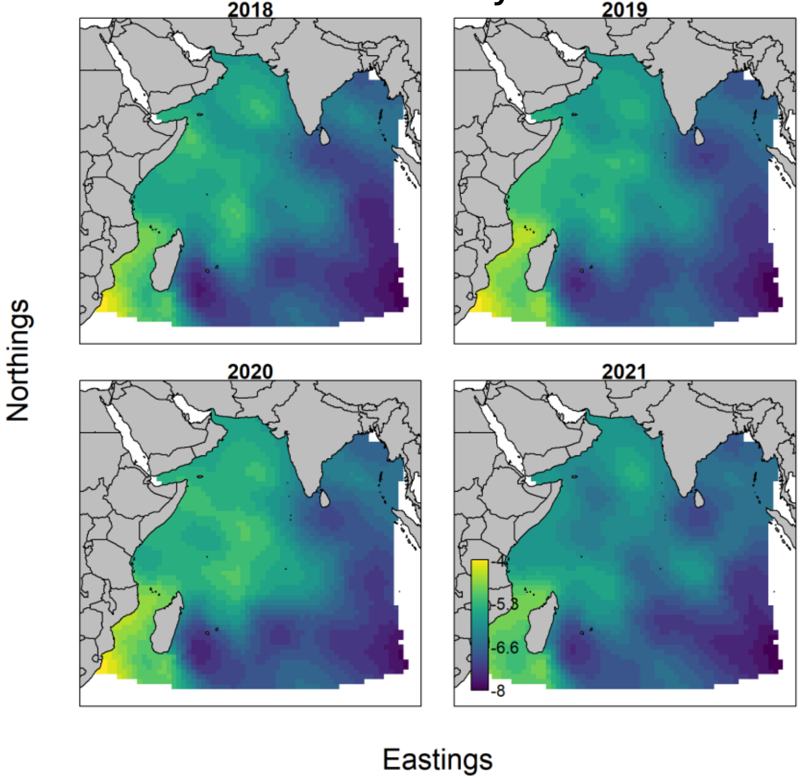
Occurrence



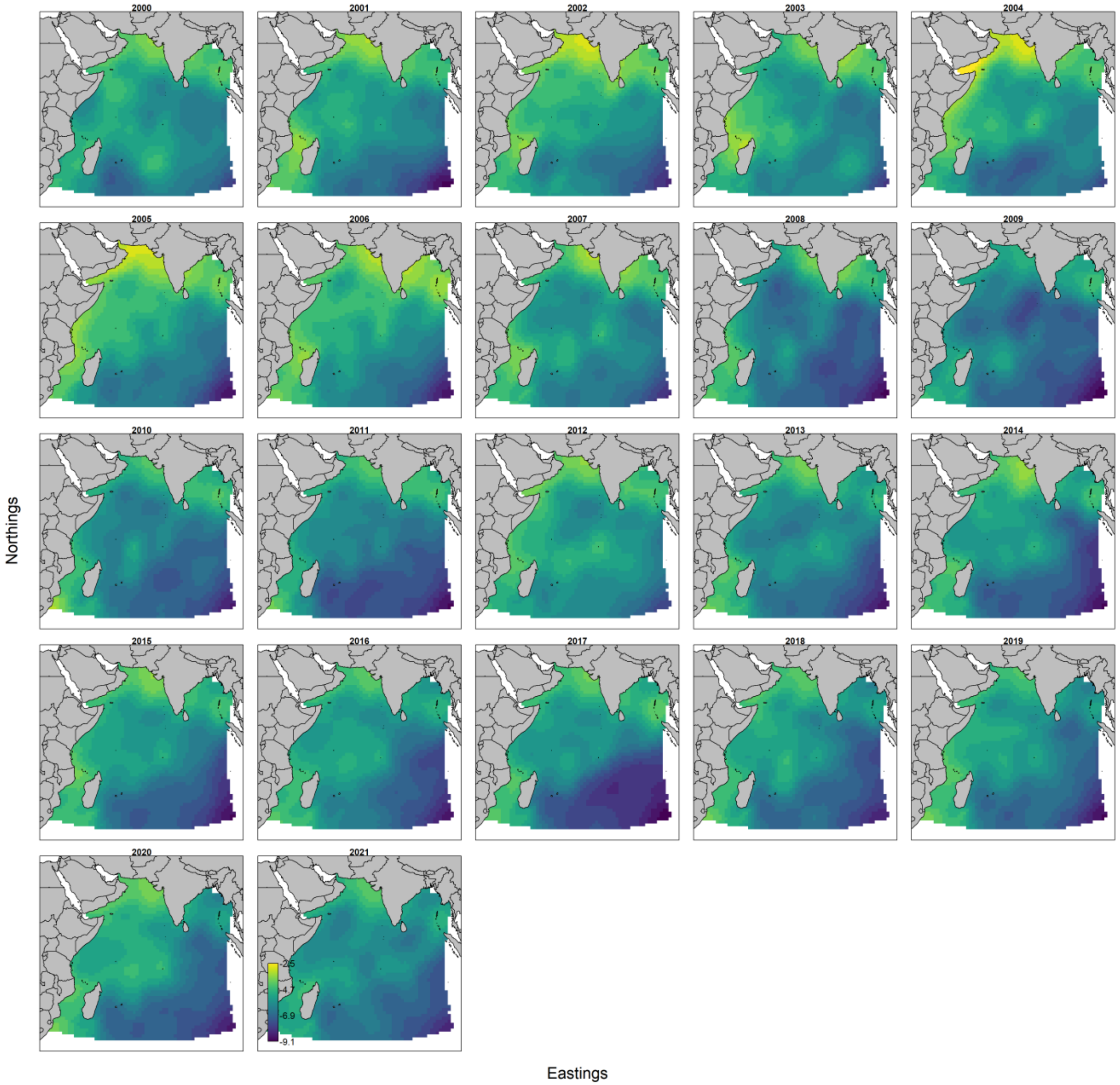
Positive catch



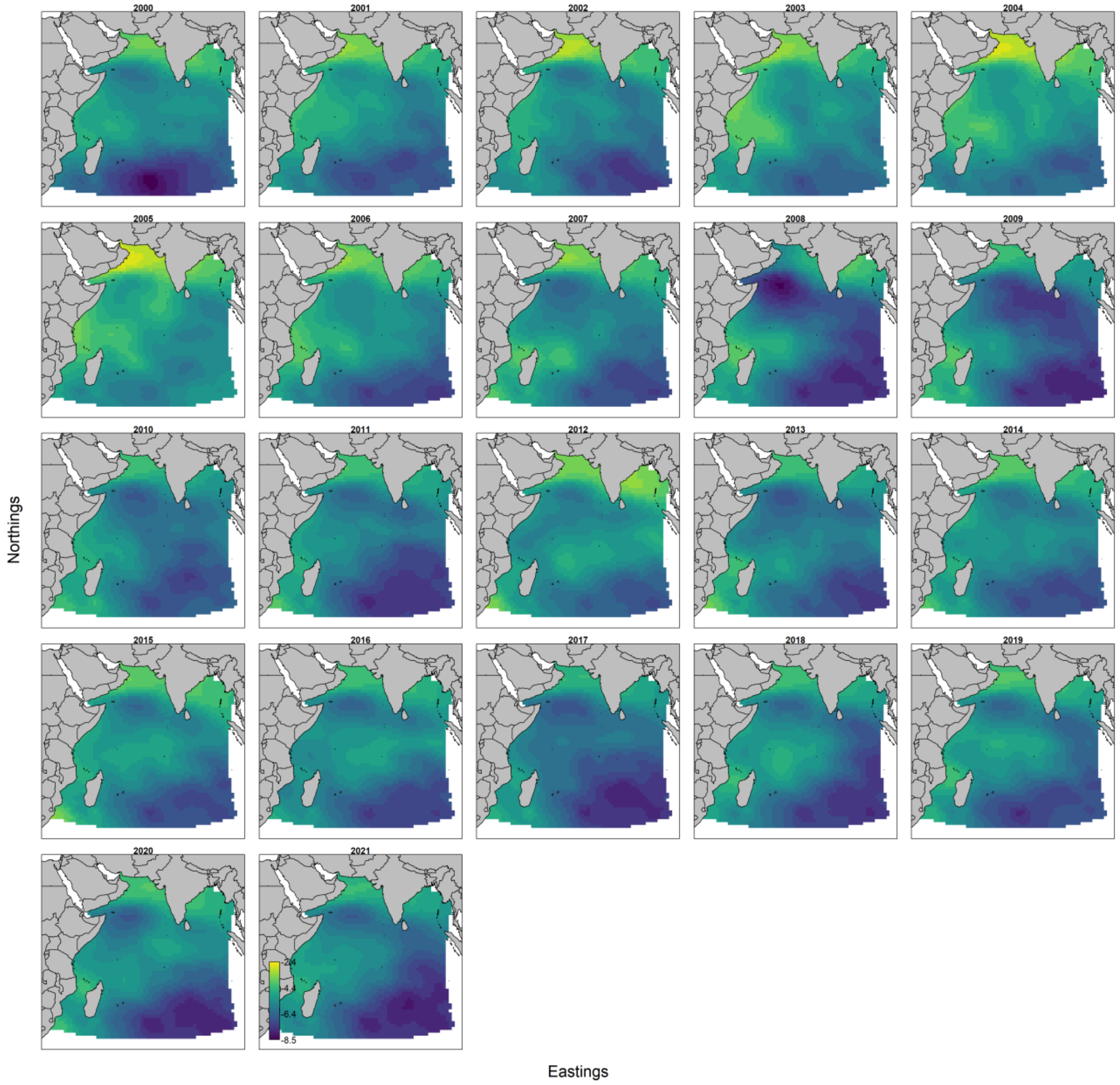
Combined
Ln Density



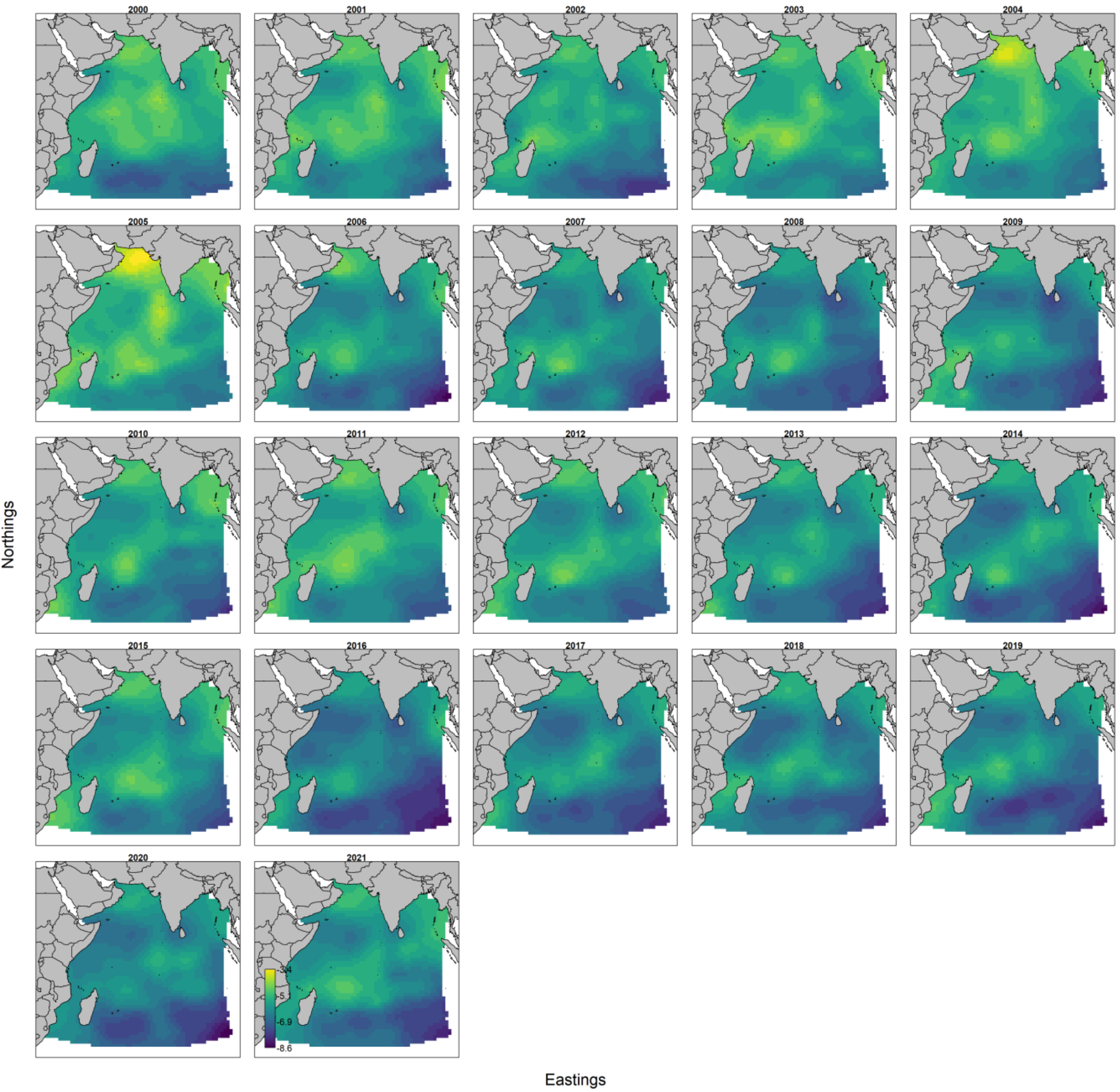
Season 1 (Jan-Mar)



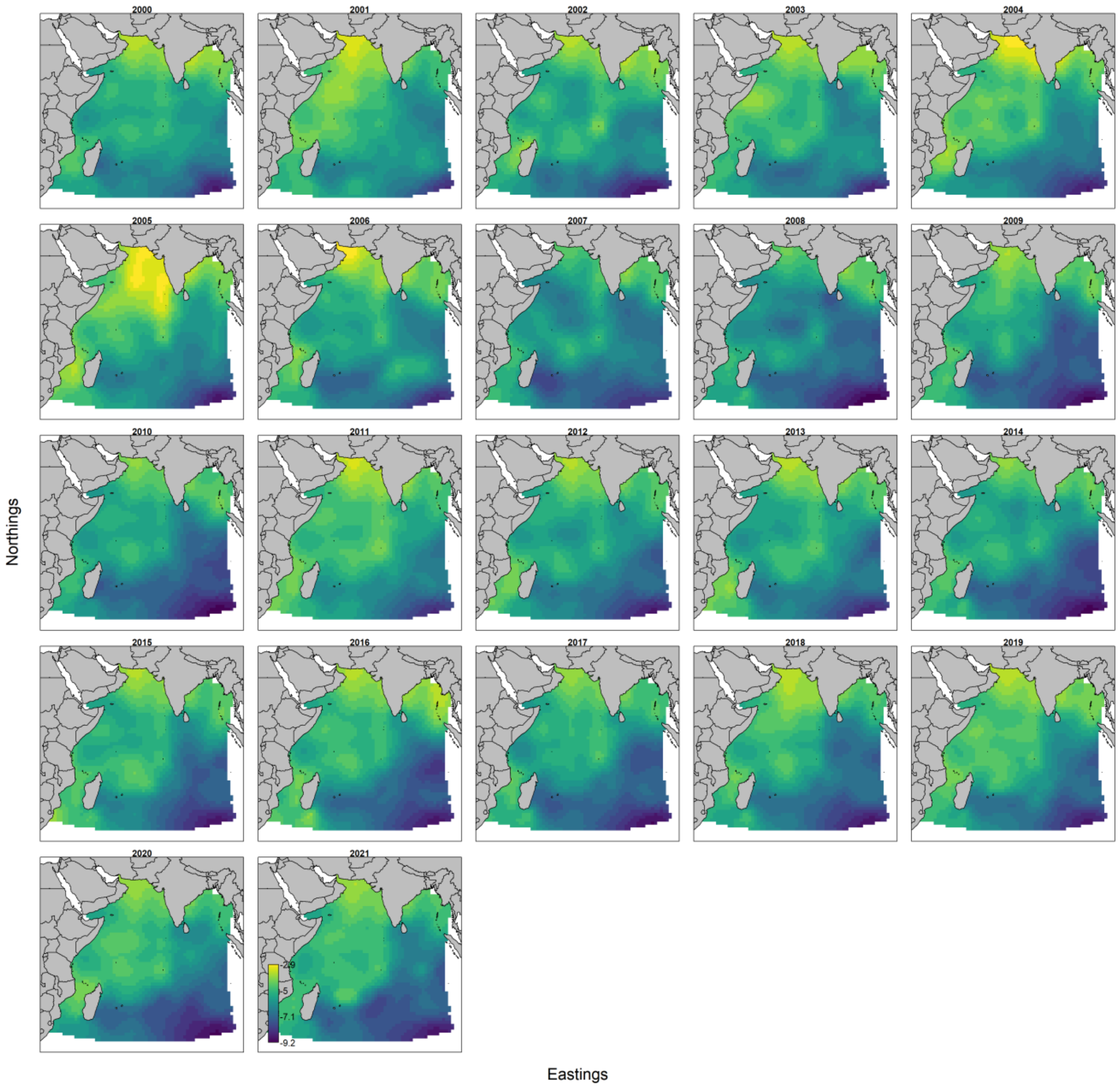
Season 2 (Apr-Jun)



Season 3 (Jul-Sep)

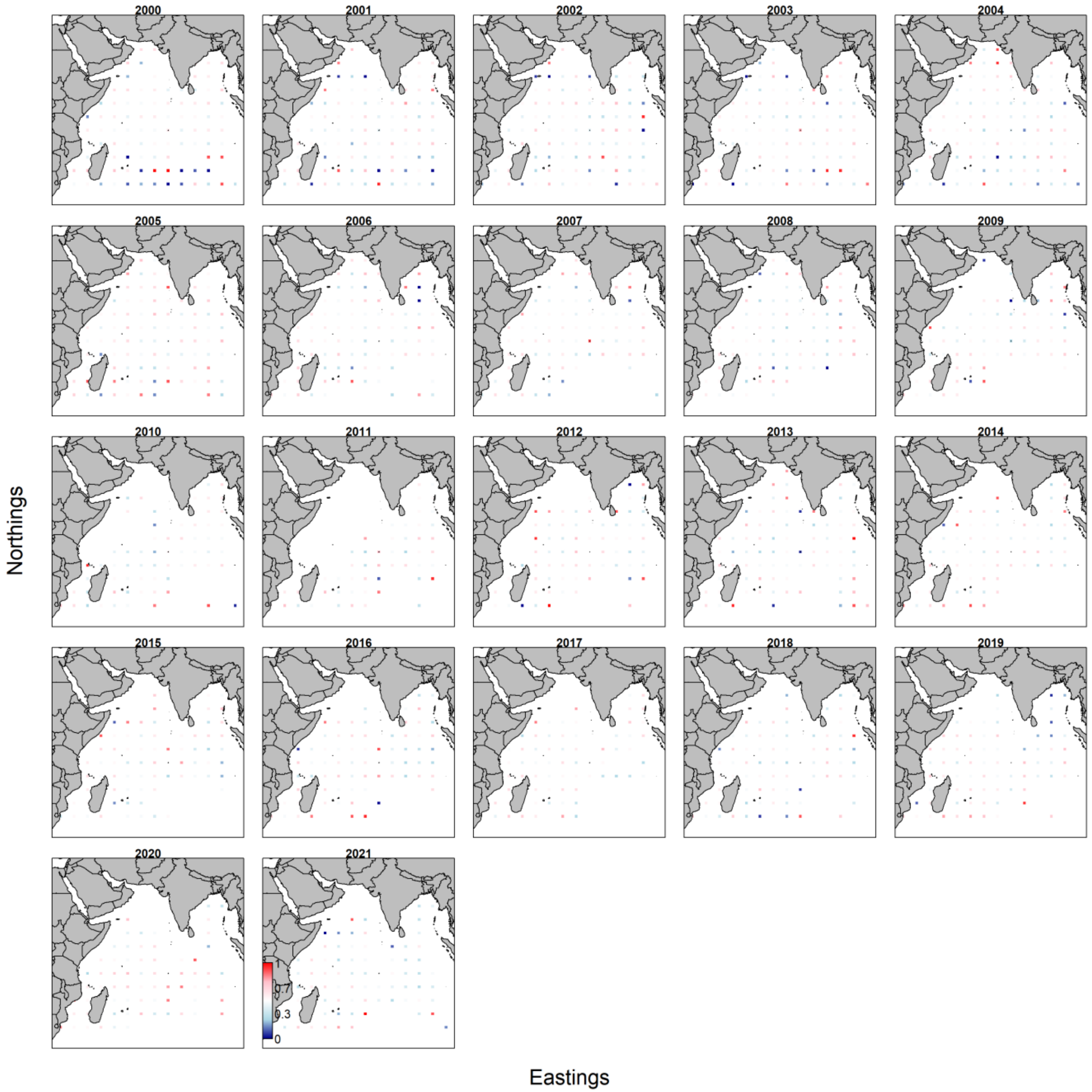


Season 4 (Oct-Dec)

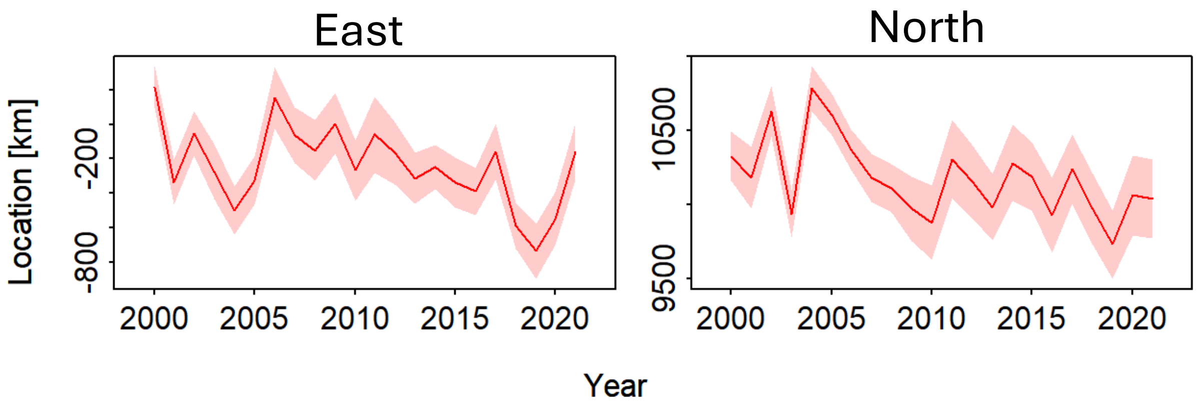


Season 1 (Jan-Mar)

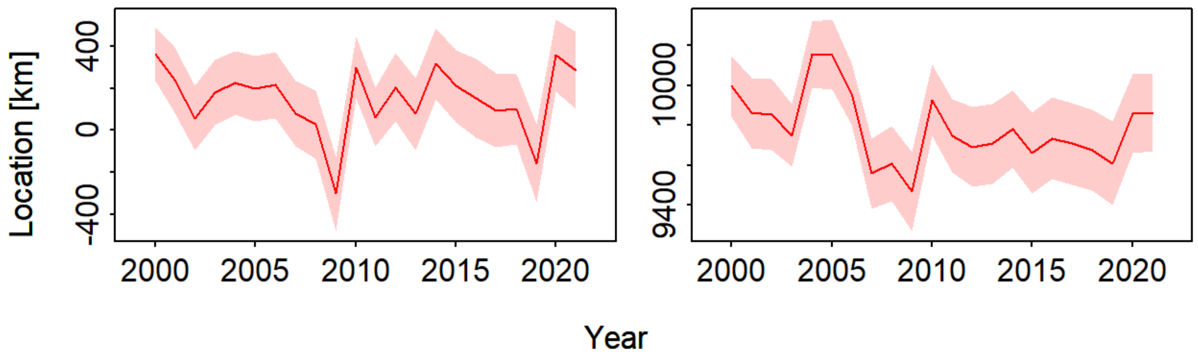
Residuals – tend to be largest in the southern area.



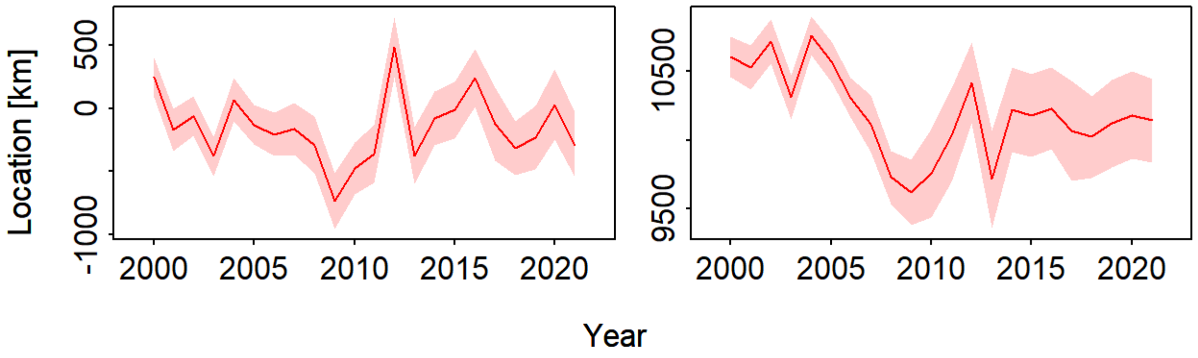
Season 1 (Jan-Mar)



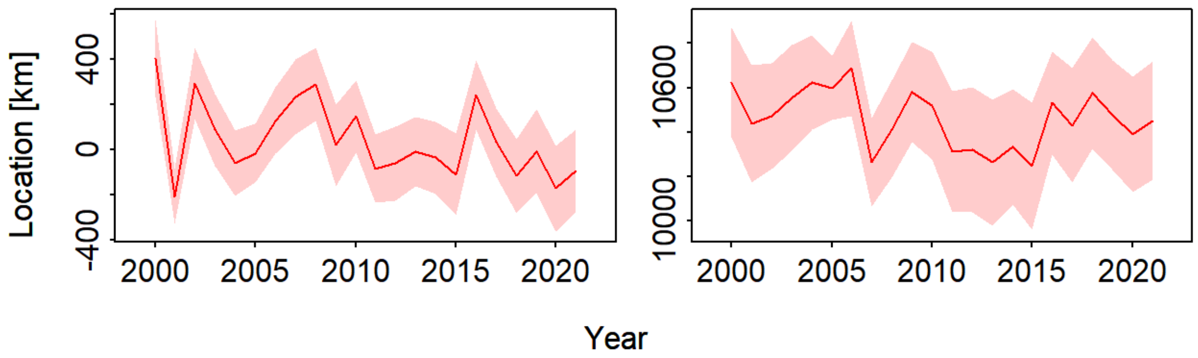
Season 3 (Jul-Sep)



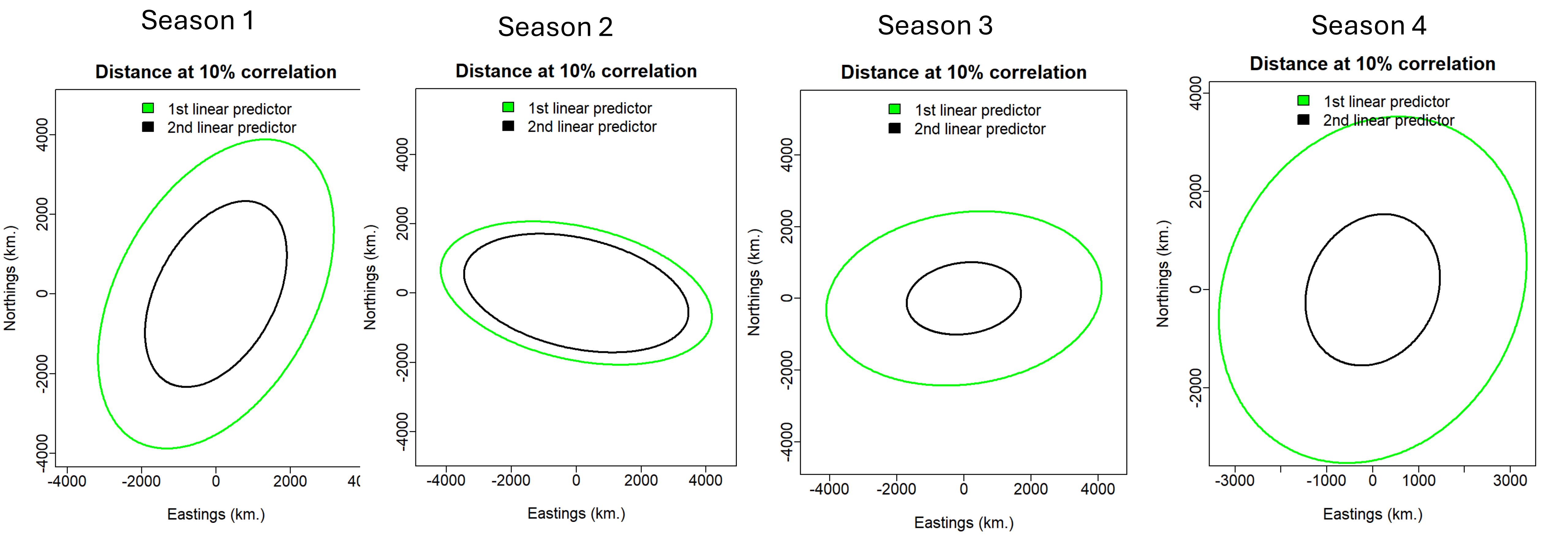
Season 2 (Apr-Jun)



Season 4 (Oct-Dec)

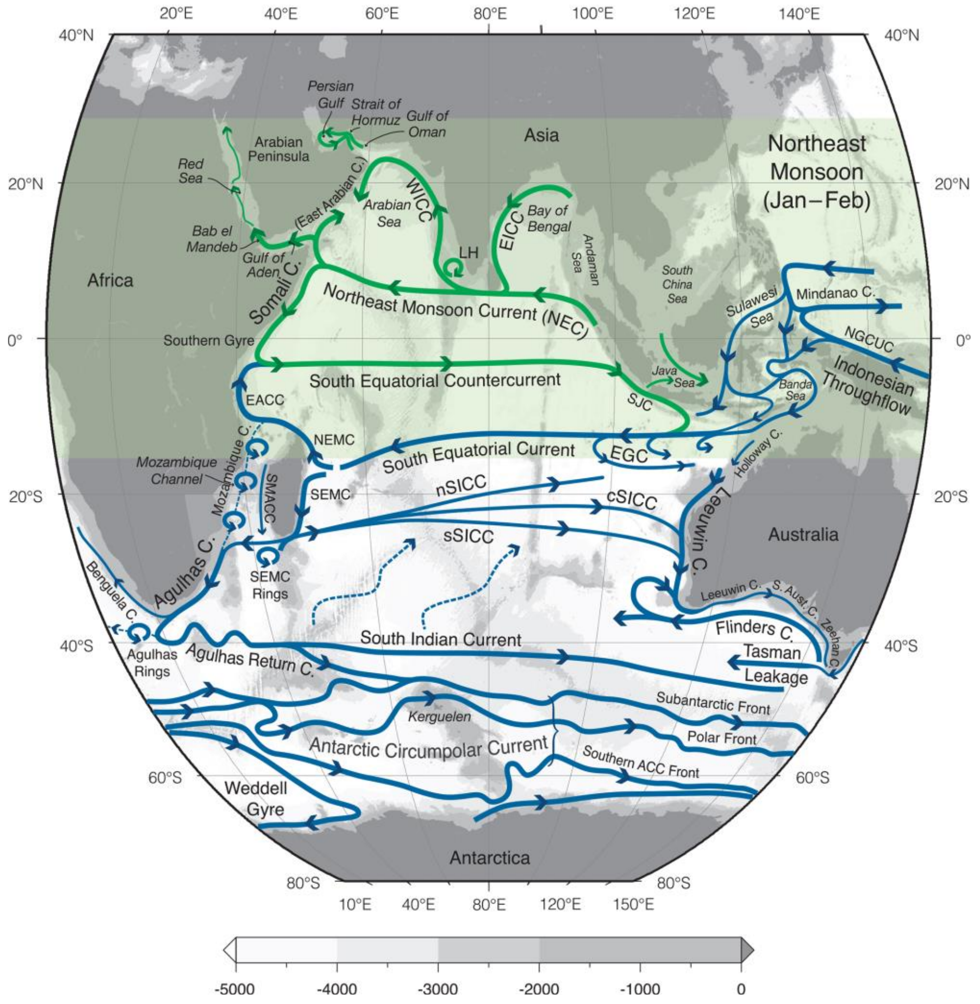


Centre of biomass (density) is further south and west in Jul-Sep (season 3).
Centre of biomass (density) is further east in Jan-Mar (season 1).
Since 2010, biomass in Jan-Mar and Oct-Dec has been further West and South (Jan-Mar).

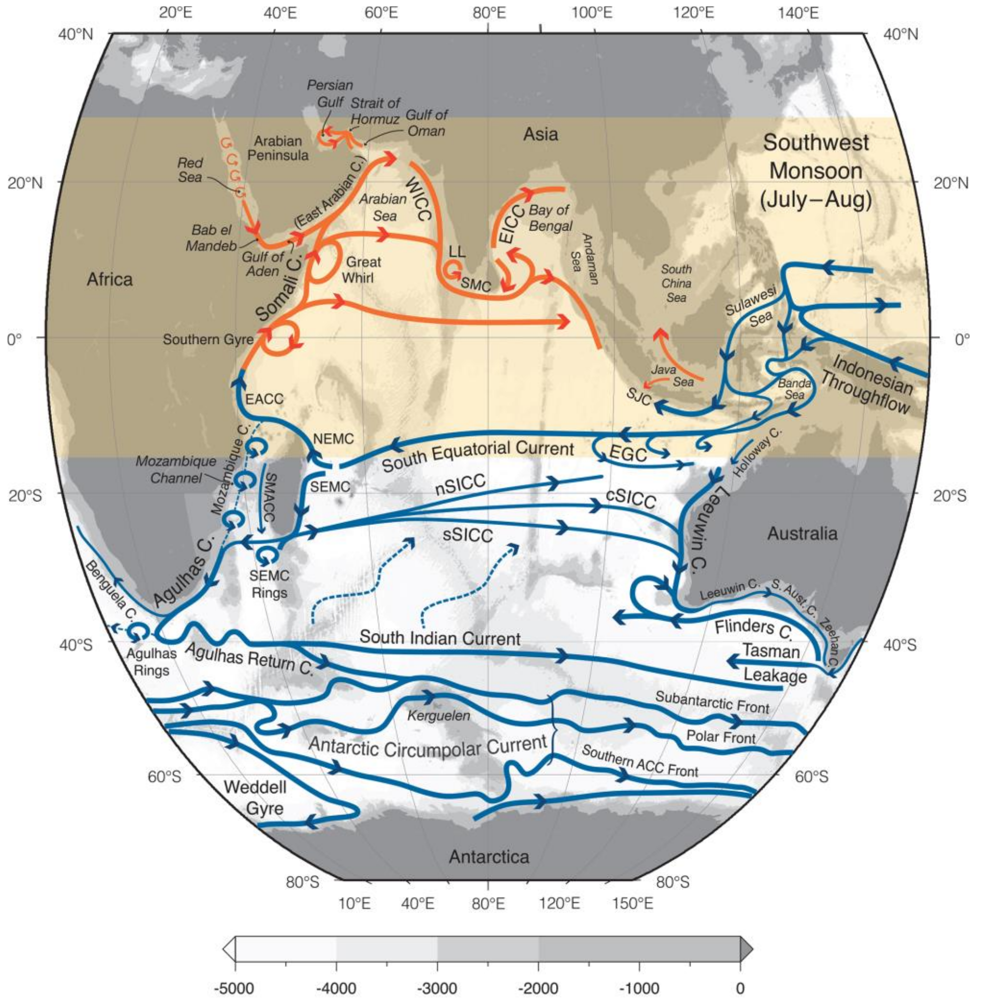


Considerable differences in spatial correlations for each season.
Probably related with the seasonal differences in circulation patterns associated with Southwest Monsoon (July-August) and Northeast Monsoon (January-February).

Season 1



Season 3



Phillips, H. E., Tandon, A., Furue, R., Hood, R., Ummenhofer, C. C., Benthuysen, J. A., Menezes, V., Hu, S., Webber, B., Sanchez-Franks, A., Cherian, D., Shroyer, E., Feng, M., Wijesekera, H., Chatterjee, A., Yu, L., Hermes, J., Murtugudde, R., Tozuka, T., Su, D., Singh, A., Centurioni, L., Prakash, S., and Wiggert, J.: Progress in understanding of Indian Ocean circulation, variability, air–sea exchange, and impacts on biogeochemistry, Ocean Sci., 17, 1677–1751, https://doi.org/10.5194/os-17-1677-2021, 2021.

Extract Region index from Season IO model

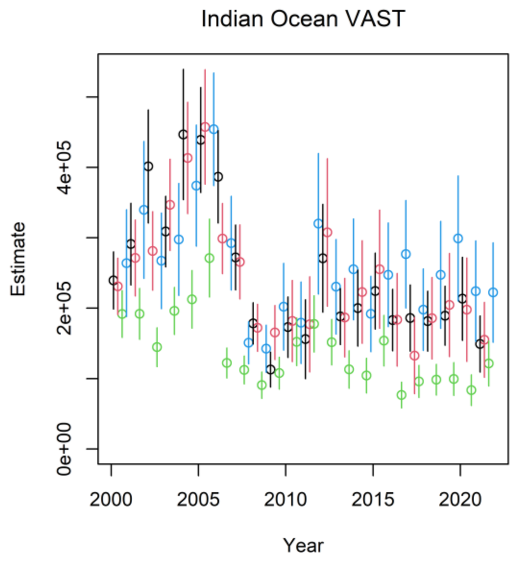
	Season 1	Season 2	Season 3	Season 4
Region 1	Yes	Yes	Yes	Yes
Region 2	Yes	Yes	Yes	Yes
Region 3	Yes	No	No	No
Region 4	Yes	No	Yes	Yes

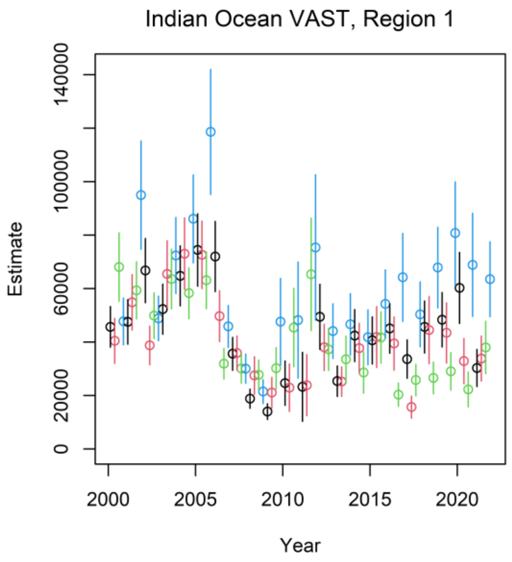
Regions treated as separate strata within Seasonal model.

No indicates convergence issues at the regional level.

Region 1

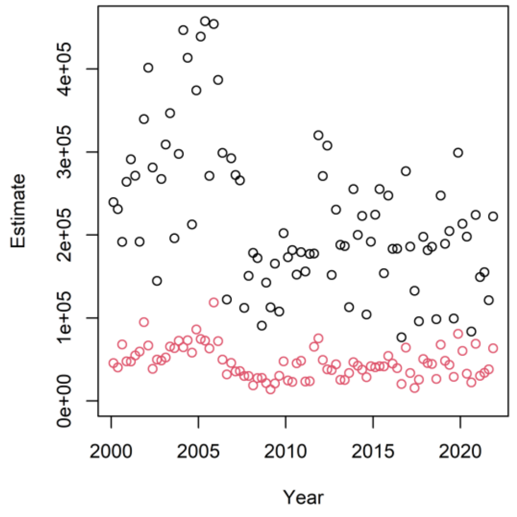
Overall Indian Ocean indices
Season 1
Season 2
Season 3
Season 4

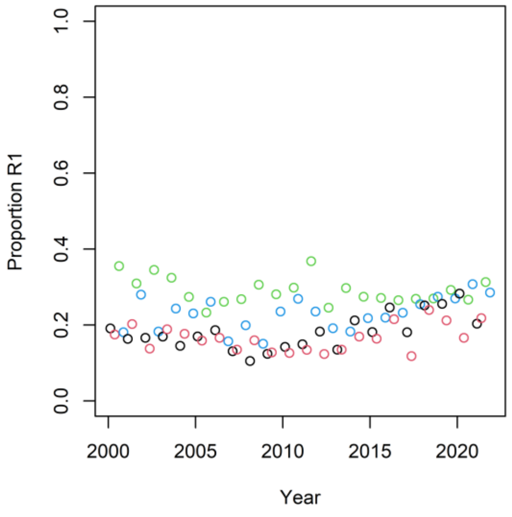




Region 1 indices (from IO VAST)
Season 1
Season 2
Season 3
Season 4

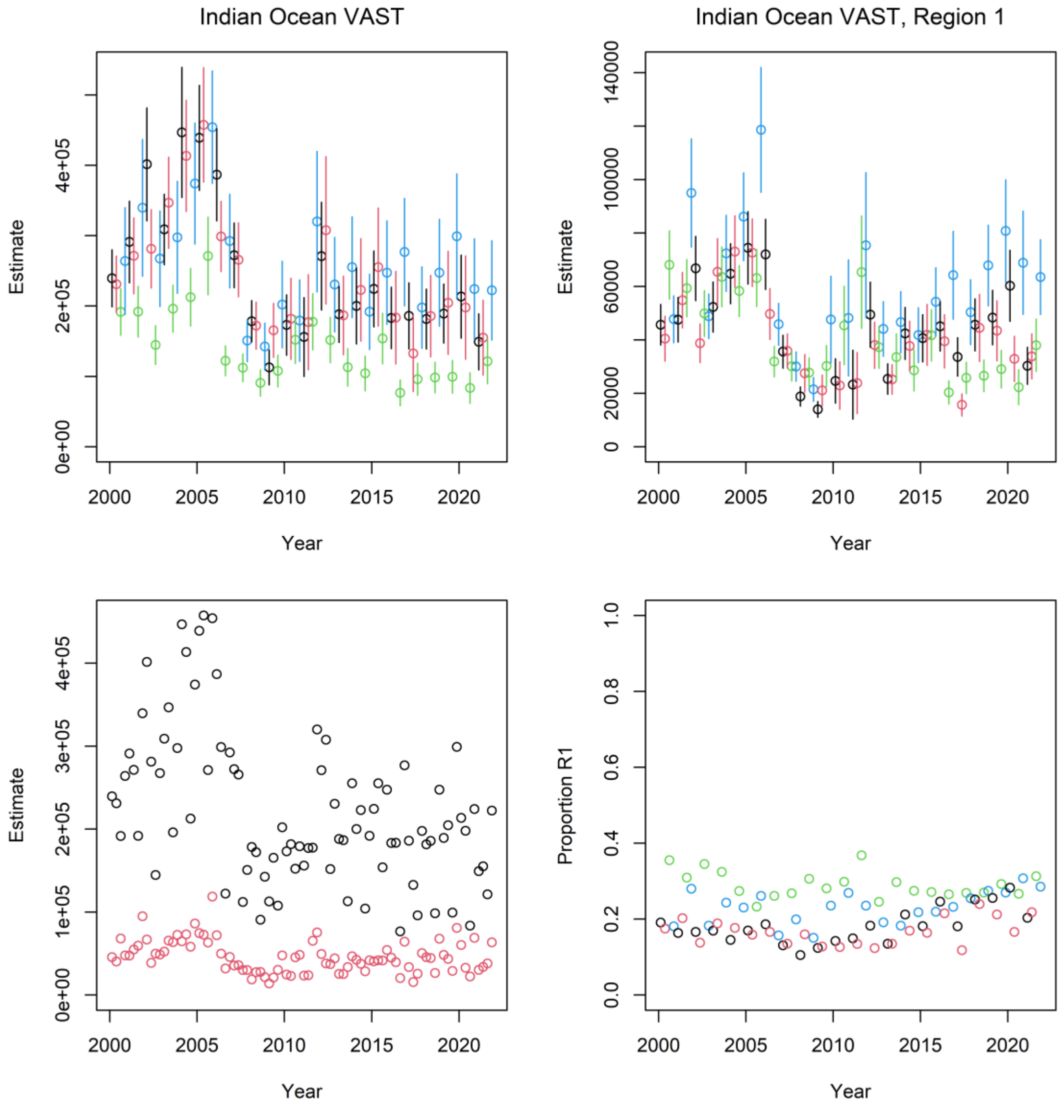
Overall Indian Ocean indices
Region 1 indices (from IO VAST)

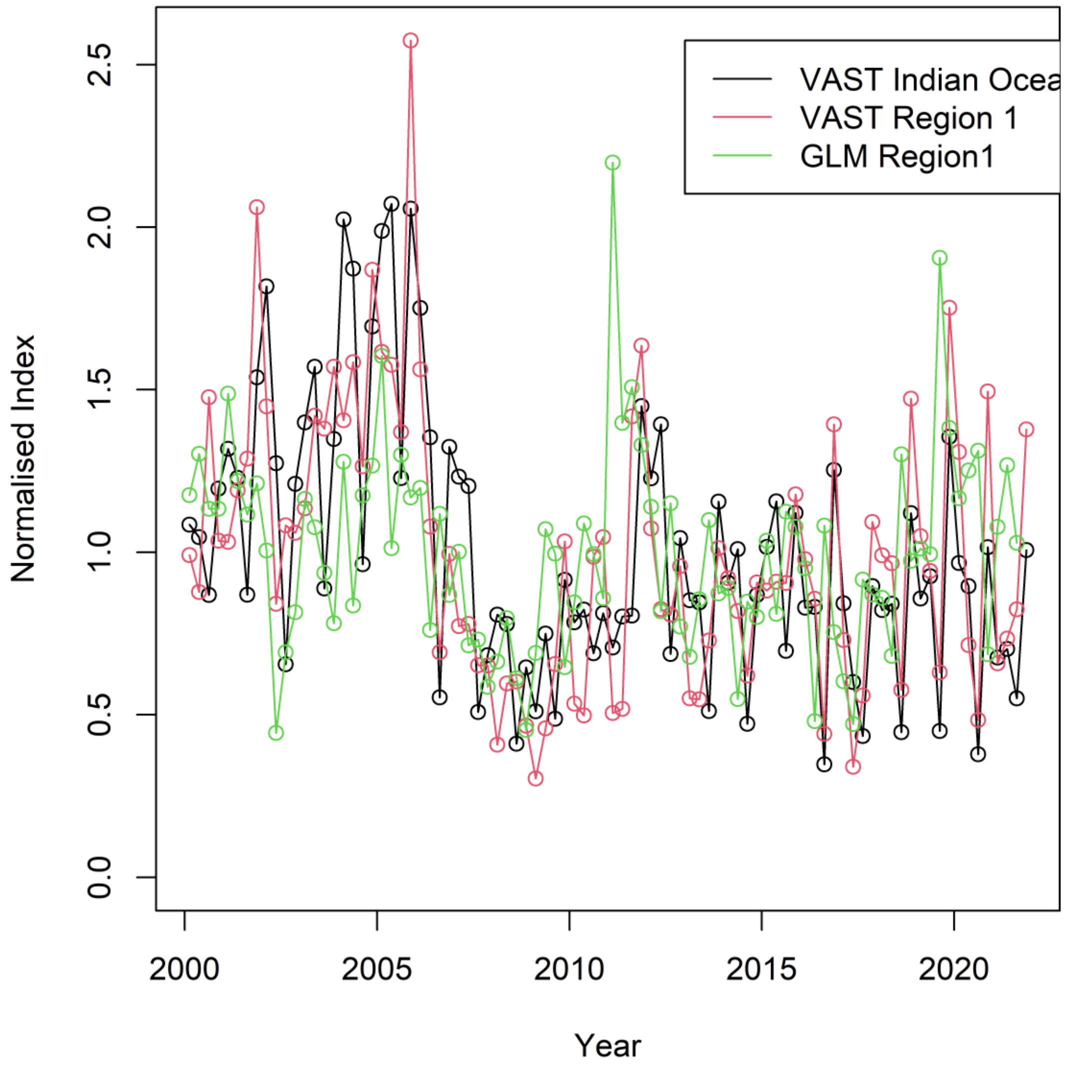




Proportion of Indian Ocean YFT within Region 1
Season 1
Season 2
Season 3
Season 4

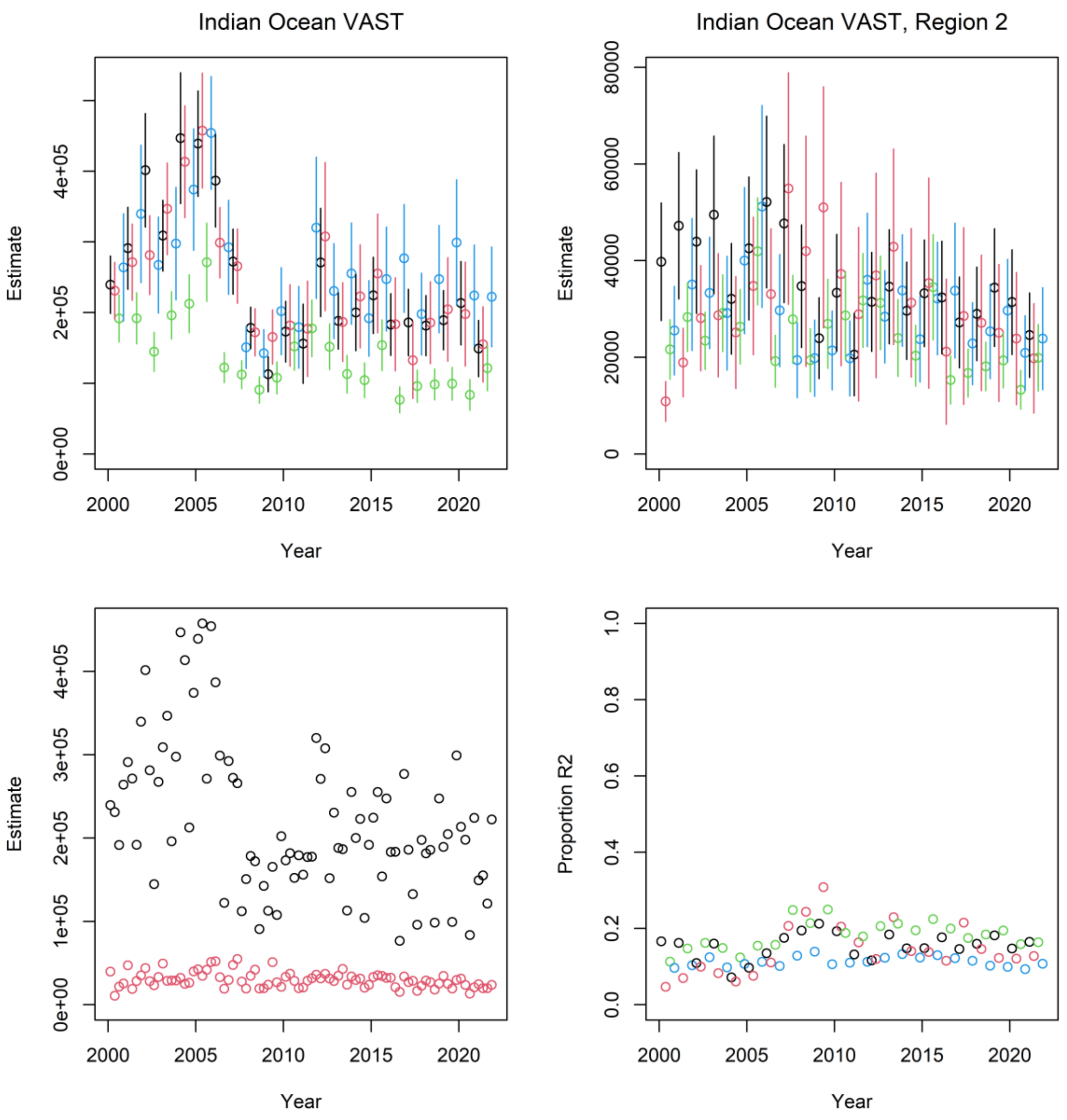
Region 1

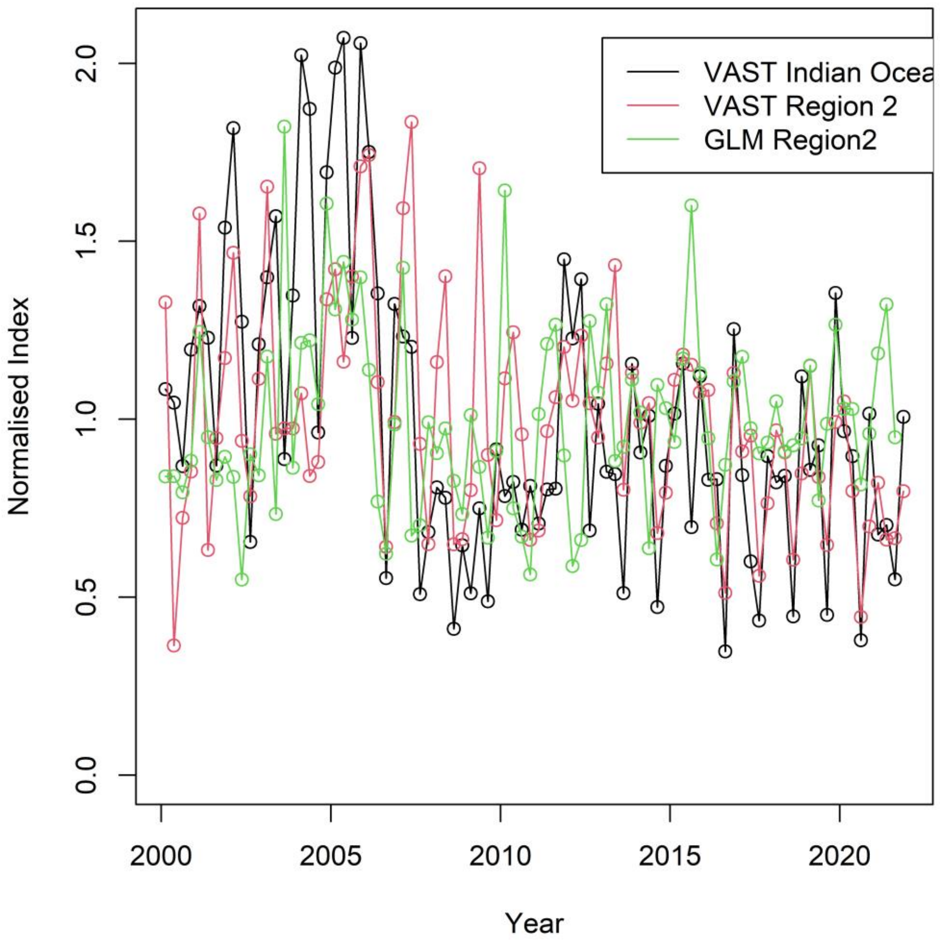




Overall IO density is lower in Season 3.
Comparable for other seasons (except higher recent S4).
Comparable GLM (this study) and VAST indices, some divergence in recent years.

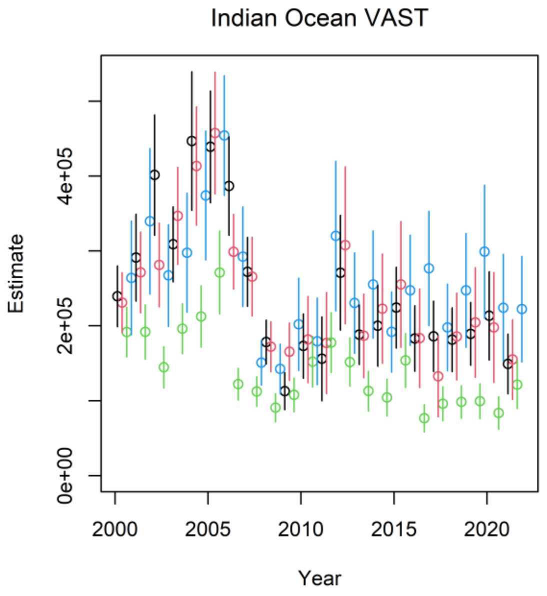
Region 2

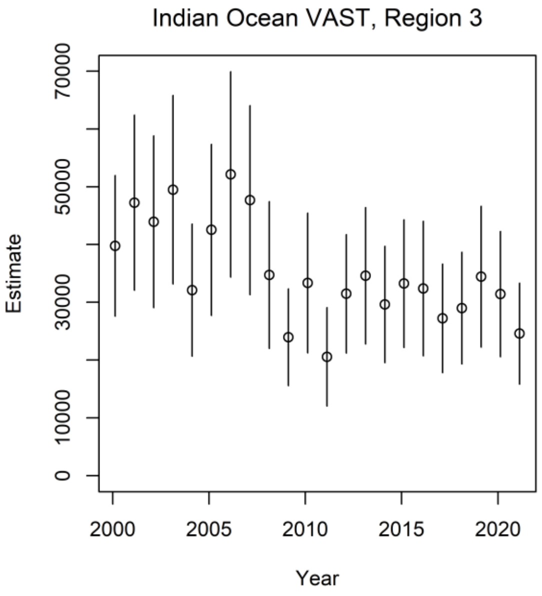


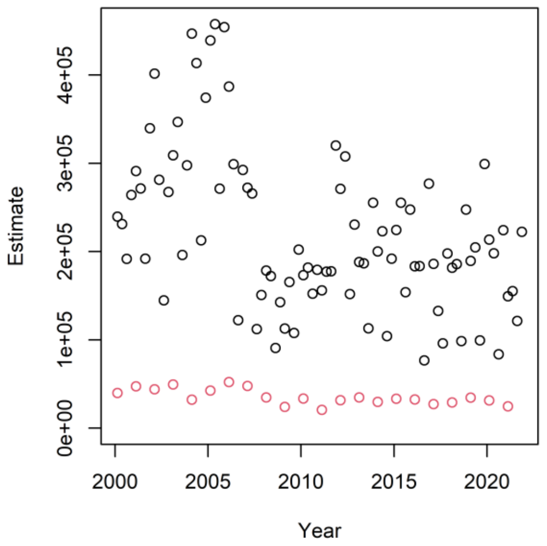


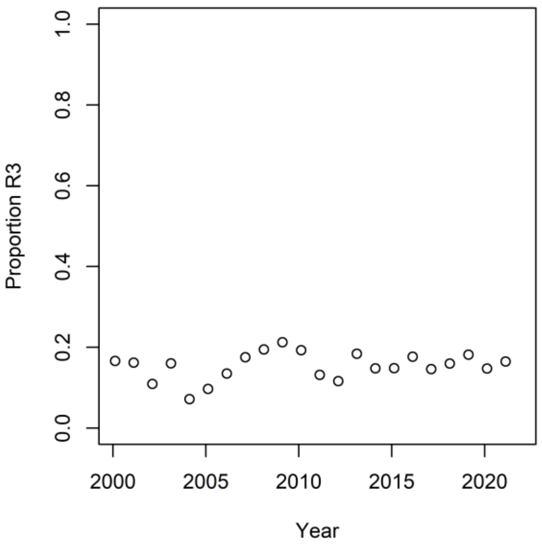
Overall IO density is lower in Season 3.
Comparable for other seasons.
Comparable GLM (this study) and VAST indices, some divergence in recent years.

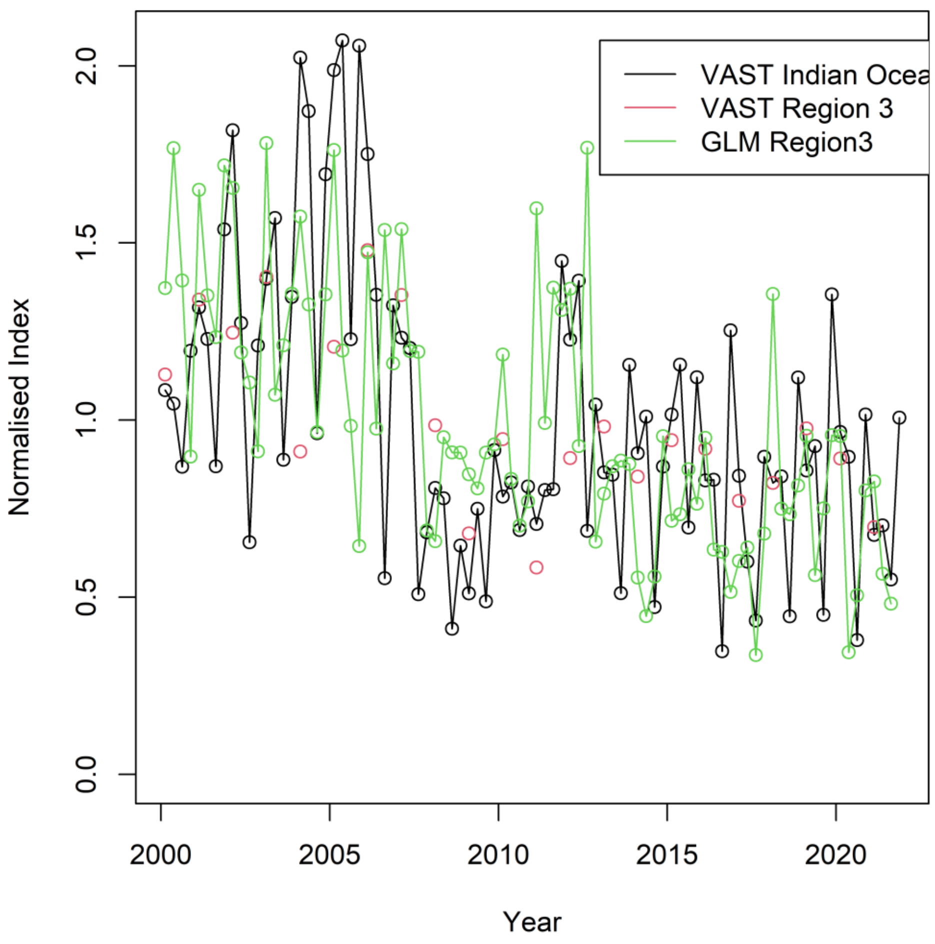
Region 3





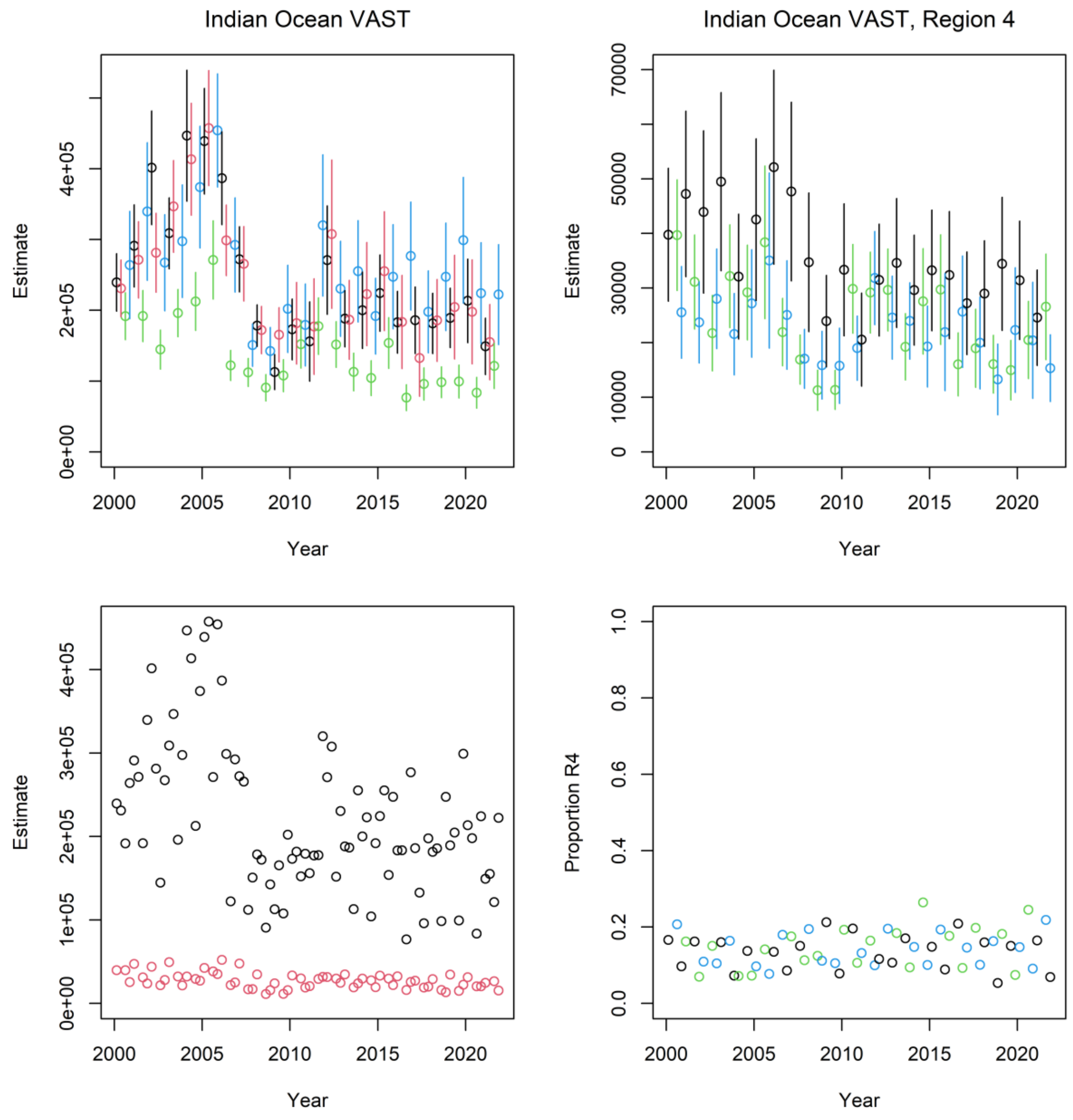


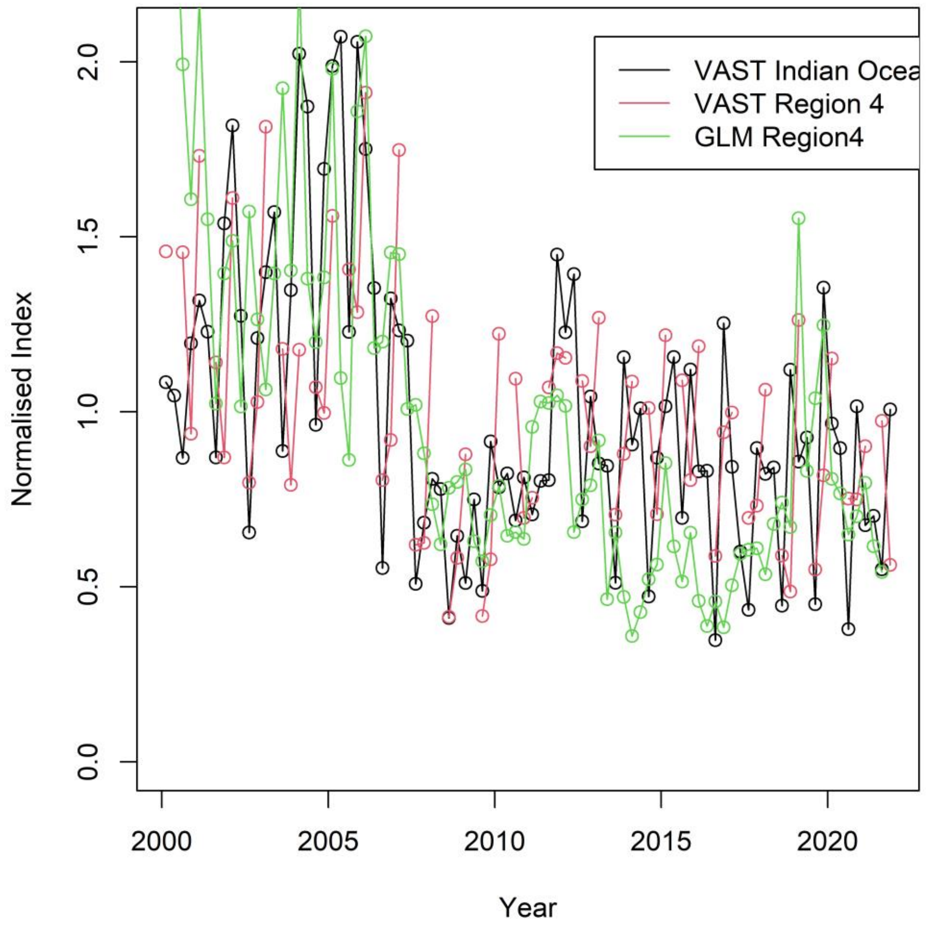




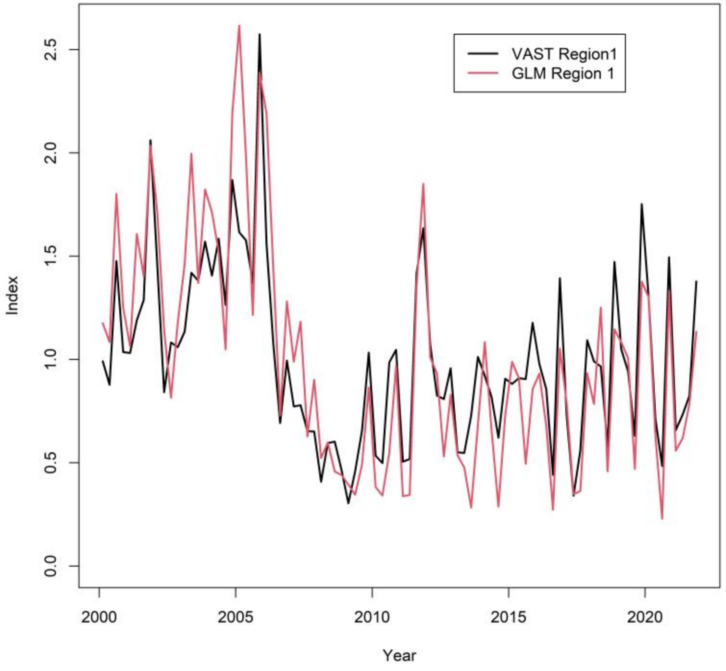
Only Season 3 VAST indices available for Region 3.

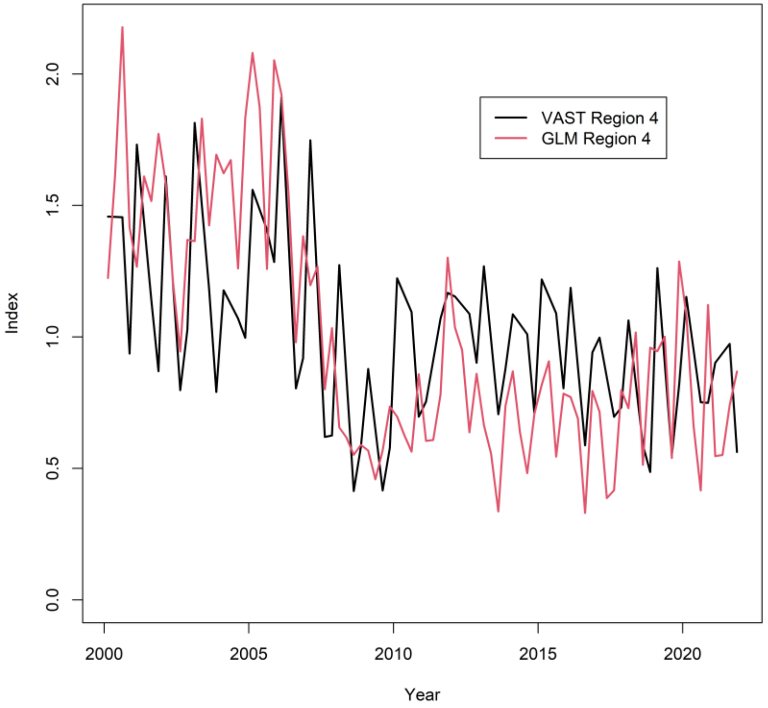
Region 4

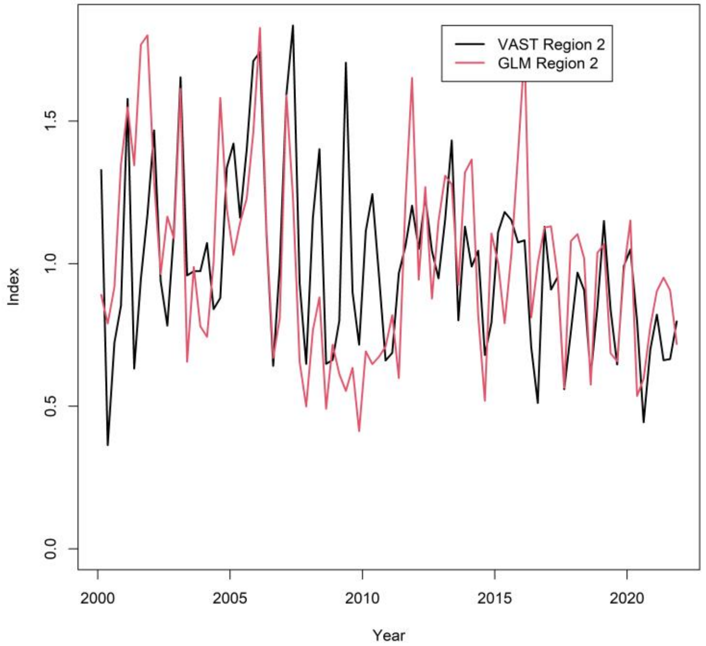




Lower Region 4 abundance in S3 & S4 (compared to S1). Broadly similar trend in R4 indices from VAST and GLM, although larger decrease in GLM indices compared to VAST indices.





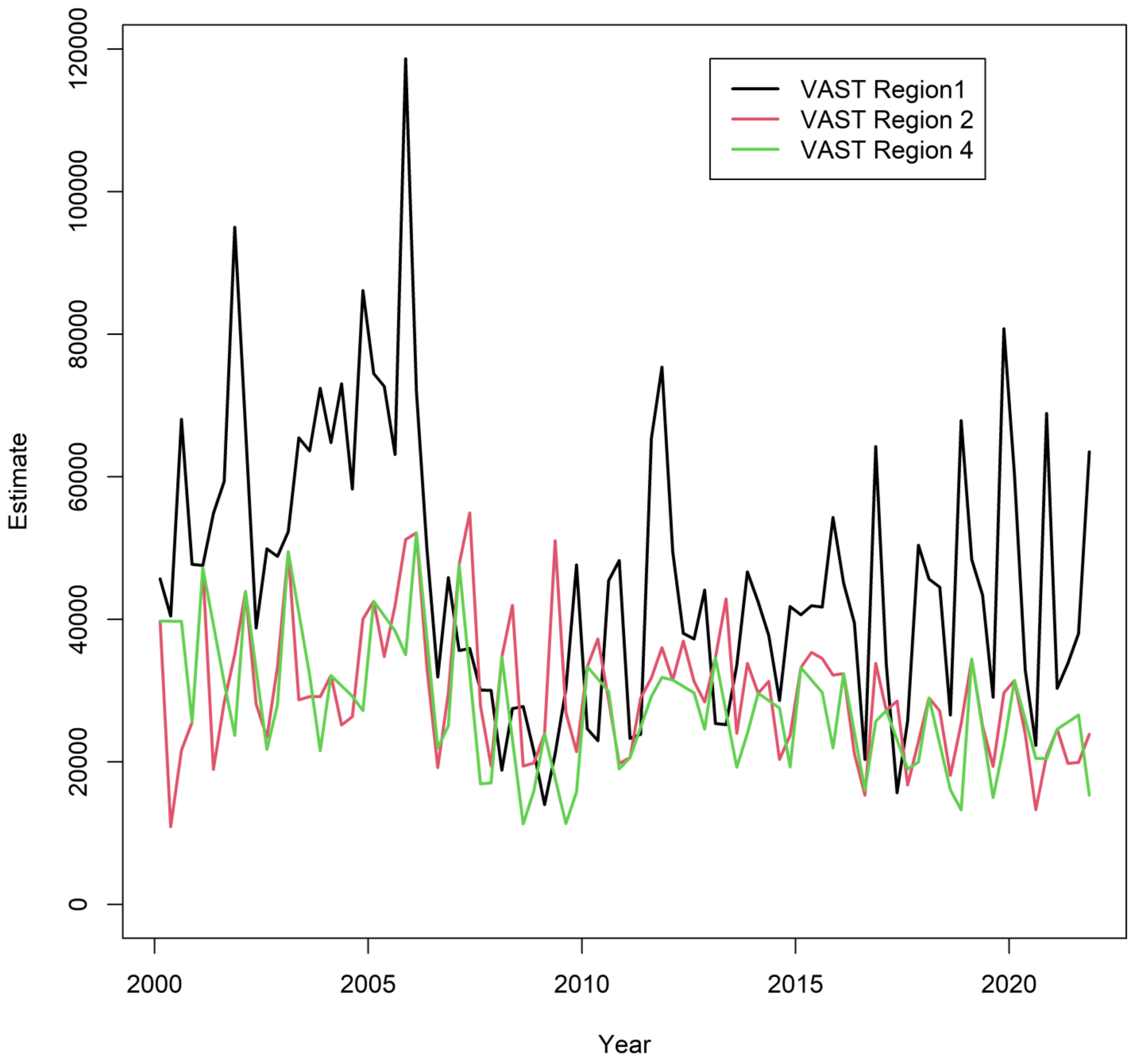


Region 1 and Region 2 VAST indices are comparable to GLM indices (except R2 late 2000s).

Larger decrease in Region 4 GLM indices compared to VAST indices.

Comparison of VAST regional indices

- Lower biomass in R2 and R4 compared to R1.
- Trends in VAST indices are comparable for R2 and R4.
- Considerably higher variation in R1 abundance.
- General decline in R2 and R4 since 2015, while R1 increasing.
- Lack of comparable indices for R3.



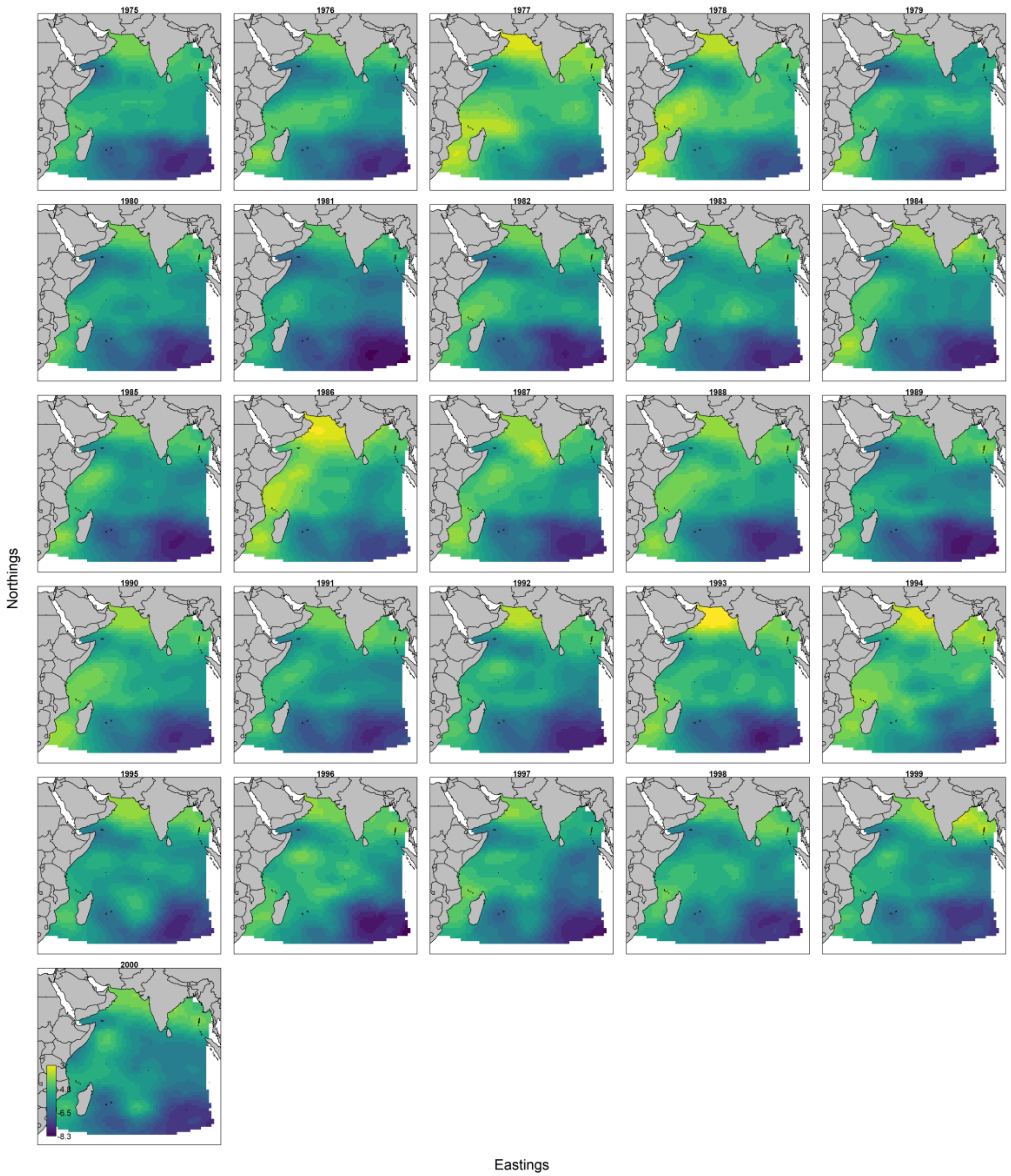
Summary 1

- VAST analysis limited to recent period. Computationally intensive.
- Limited set of predictor variables (Nation). No HBF data or species comp clustering.
- Overall spatial effects vary with season (esp. S3).
- Indication of spatial shift in YFT distribution since 2010 (but also coincides with a change in the distribution of the fishing records).
- Lower density (catchability) during Season 3. Also limited data and convergence issues.
- Some divergence in regional CPUE trends between Seasons, especially in recent years. Comparable trends R2 and R4, deviate from R1.
- Broadly consistent with region specific GLM indices, some exceptions: VAST R2 less optimistic than GLM, R4 more optimistic.
- VAST analysis indicates regional distribution of biomass: R1 20-30% (excluding Arabian Sea), R2 10-20%, R3 20%, R4 15-20%. Compared to current assessment model regional scaling of R1 45%, R2 17%, R3 12%, R4 27% (derived from different time periods).
- Limited data coverage in recent years, e.g. Somali EEZ in recent years as well as 2007-2010 (piracy).
- Potential to incorporate catchability covariates (e.g. SST, mixed layer depth) to inform spatial predictions.

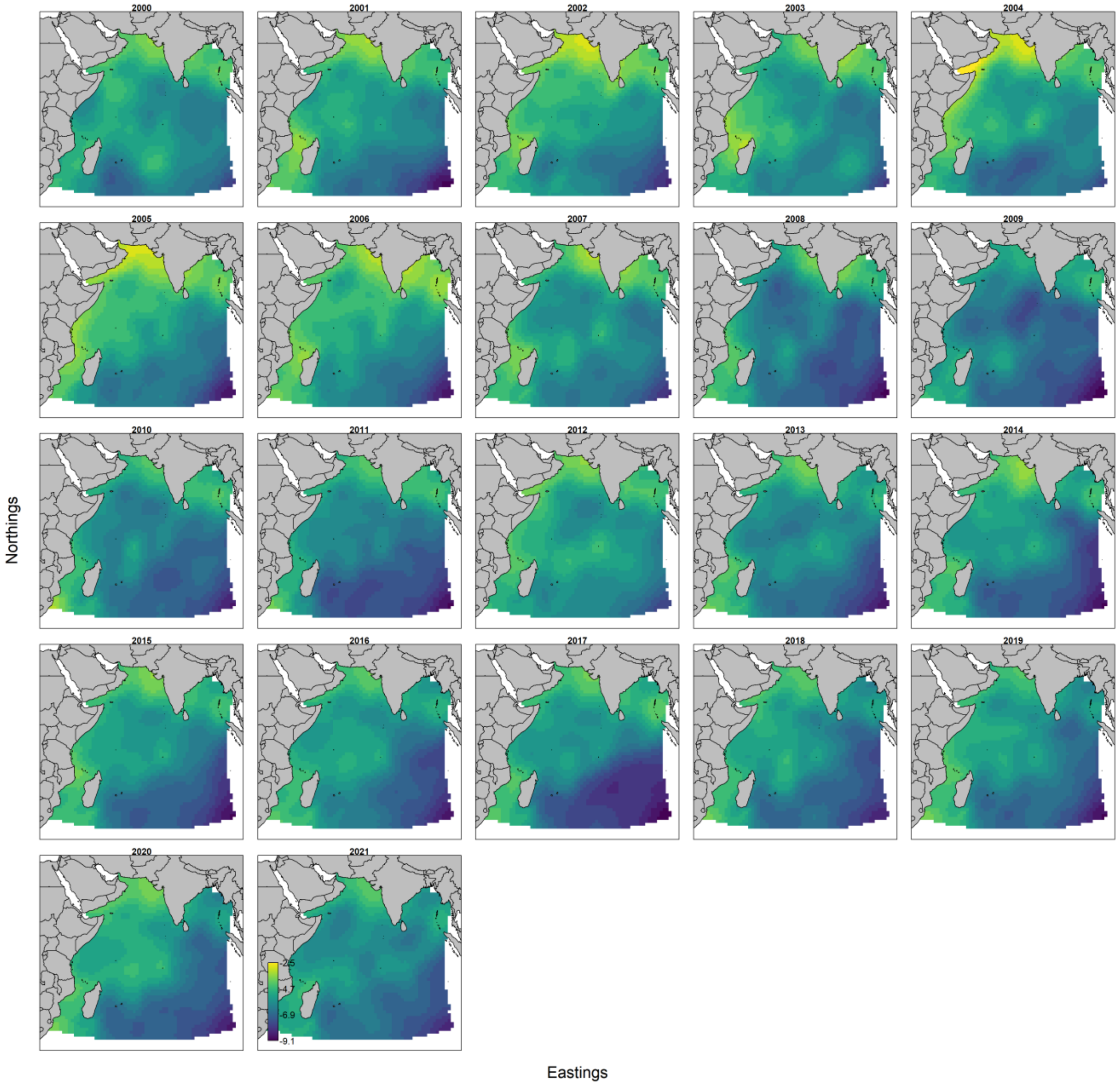
Early period

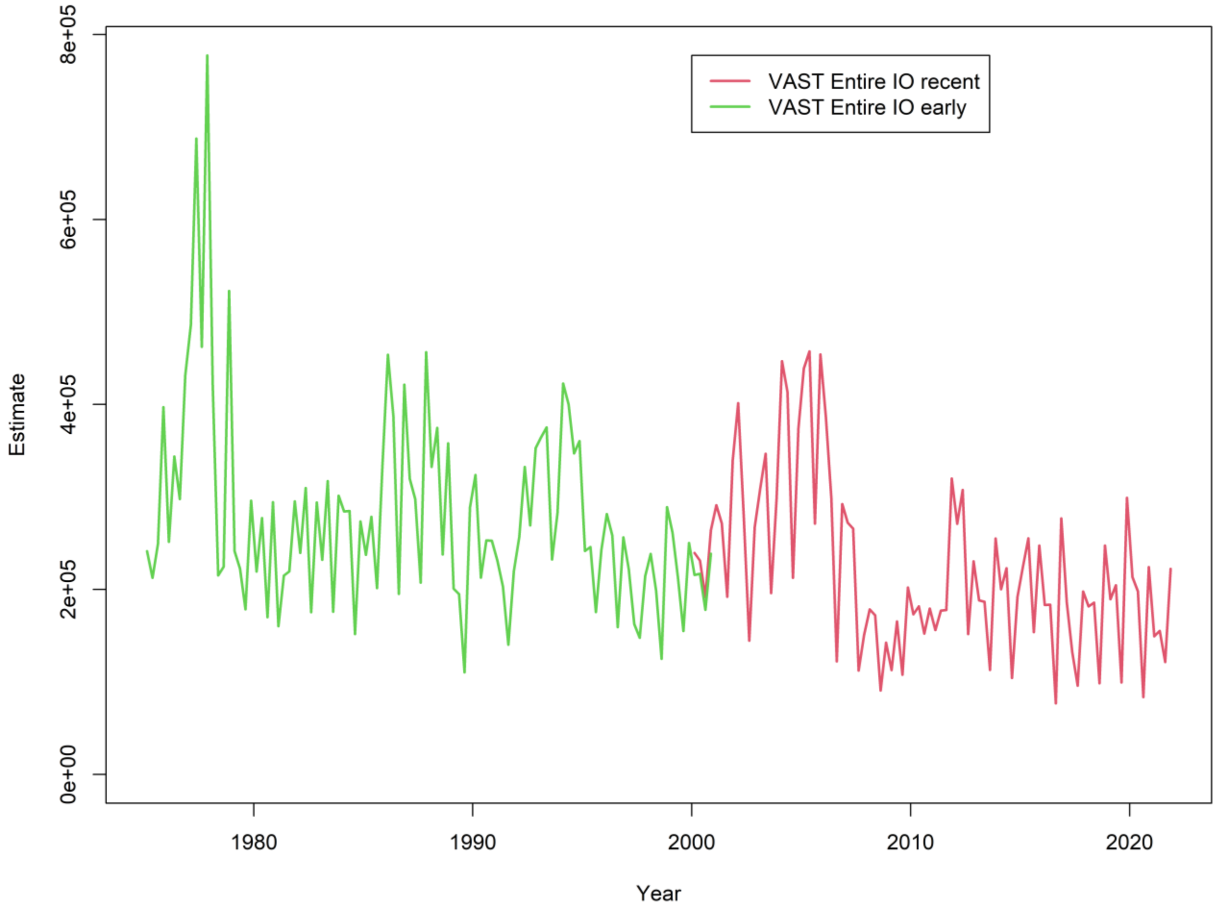
- Repeat Base model VAST analysis for 1975-2000 (4 seasonal models).
- Comparable level of “biomass” for entire Indian Ocean relative to “recent” period.
- Fluctuating “biomass” during 1980-2008, lower biomass since 2010.
- Extrapolation into Arabian Sea (and Bay of Bengal) during periods with little/no data. See data coverage 1975-1985, 2004-2021.
- Similar long-term trend to comparable GLM indices.

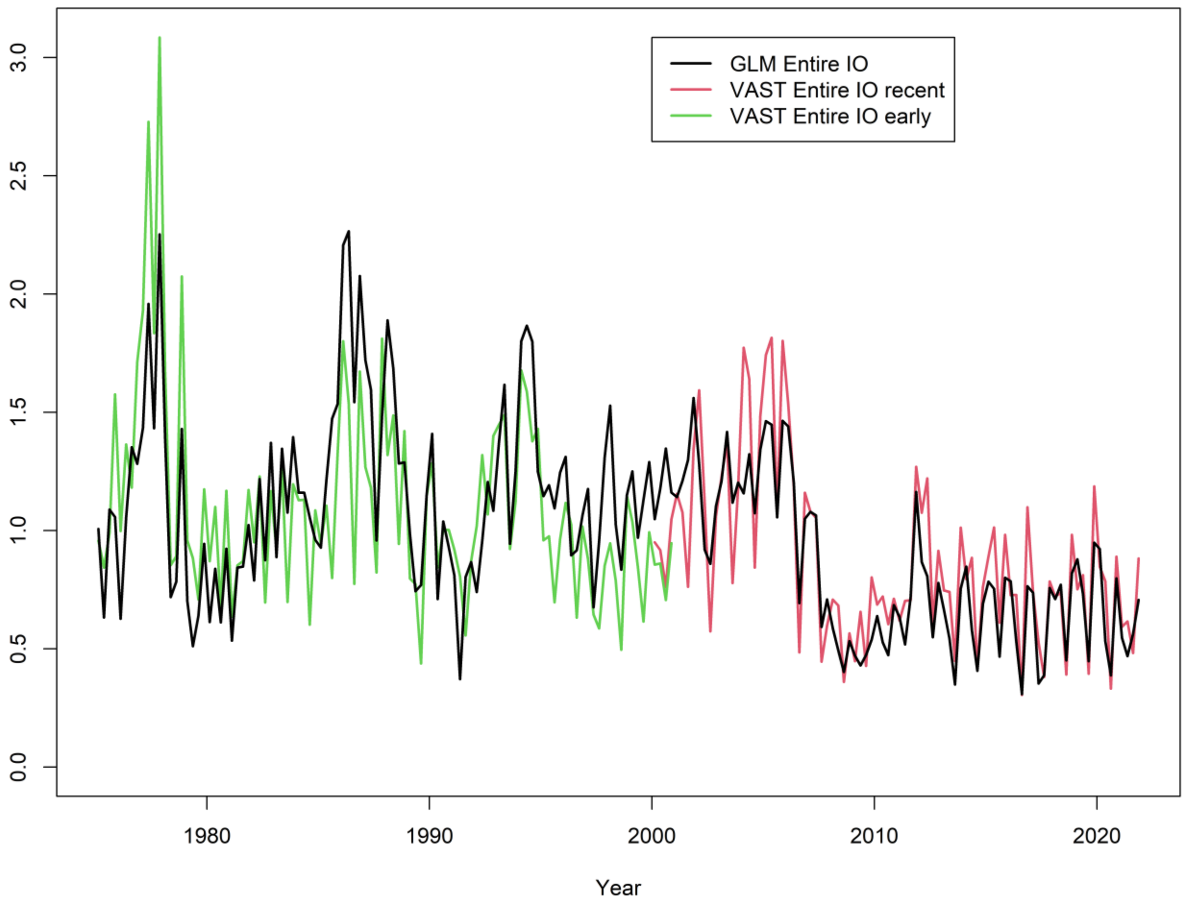
Season 1, 1975-2000



Season 1, 2000-2021




VAST Entire IO recent
VAST Entire IO early
Estimate
0e+00
2e+05
4e+05
6e+05
8e+05
1980
1990
2000
2010
2020
Year


GLM Entire IO
VAST Entire IO recent
VAST Entire IO early
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1980
1990
2000
2010
2020
Year

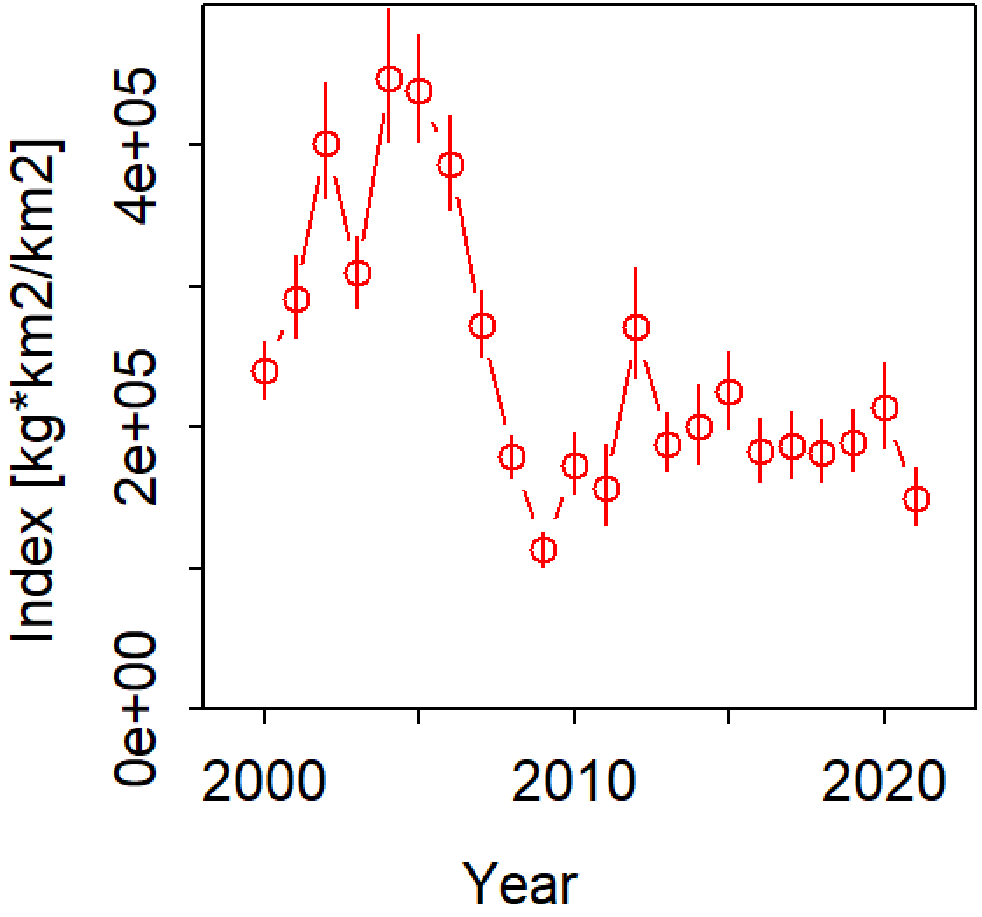
Spatial Temporal Effects, incorporating Environmental Covariates (Season 1, 2000-2021)

Model + covariate	**Obj Function**	**cAIC**	**No. coefficients (Random Effects)**
Base model (Season 1) (include SpatialTemporal random effects)	39742	79591	26694 (26640)
Base + MLD	39741	79595	26720 (26664)
Base + SST	39741	79595	26720 (26664)
Base + CurrentU	39754	79617	26719 (26664)
Base_exST (exclude SpatialTemporal random effects)	40229	80559	1161 (1111)
Base_exST + CurrentU + SST + MLD (exclude SpatialTemporal random effects)	40179	80474	2280 (2222)

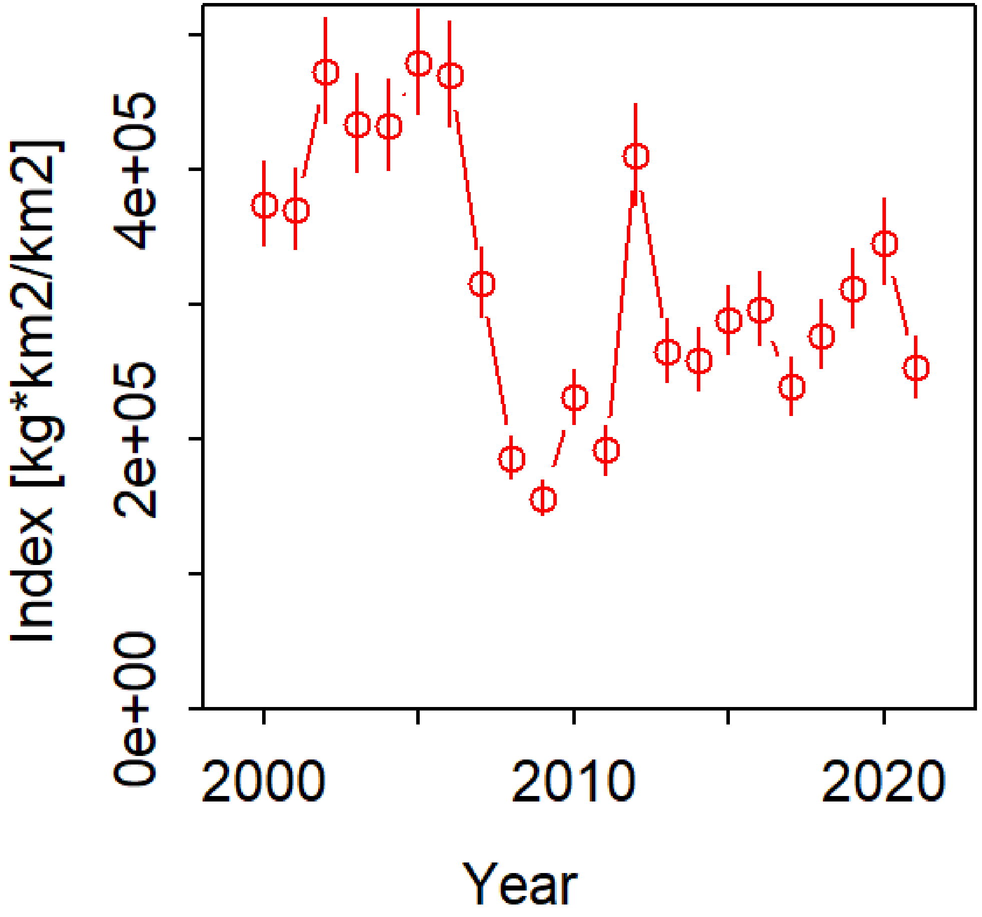
SST, MLD, CurrentU (East-West) sourced from NCEP.

Covariates included as 2nd order polynomial functions.
X2_formula = ~ poly(sst, degree=2) + poly(mld, degree=2) + poly(currentu, degree=2)

Base model (Season 1)
(include SpatialTemporal random effects)

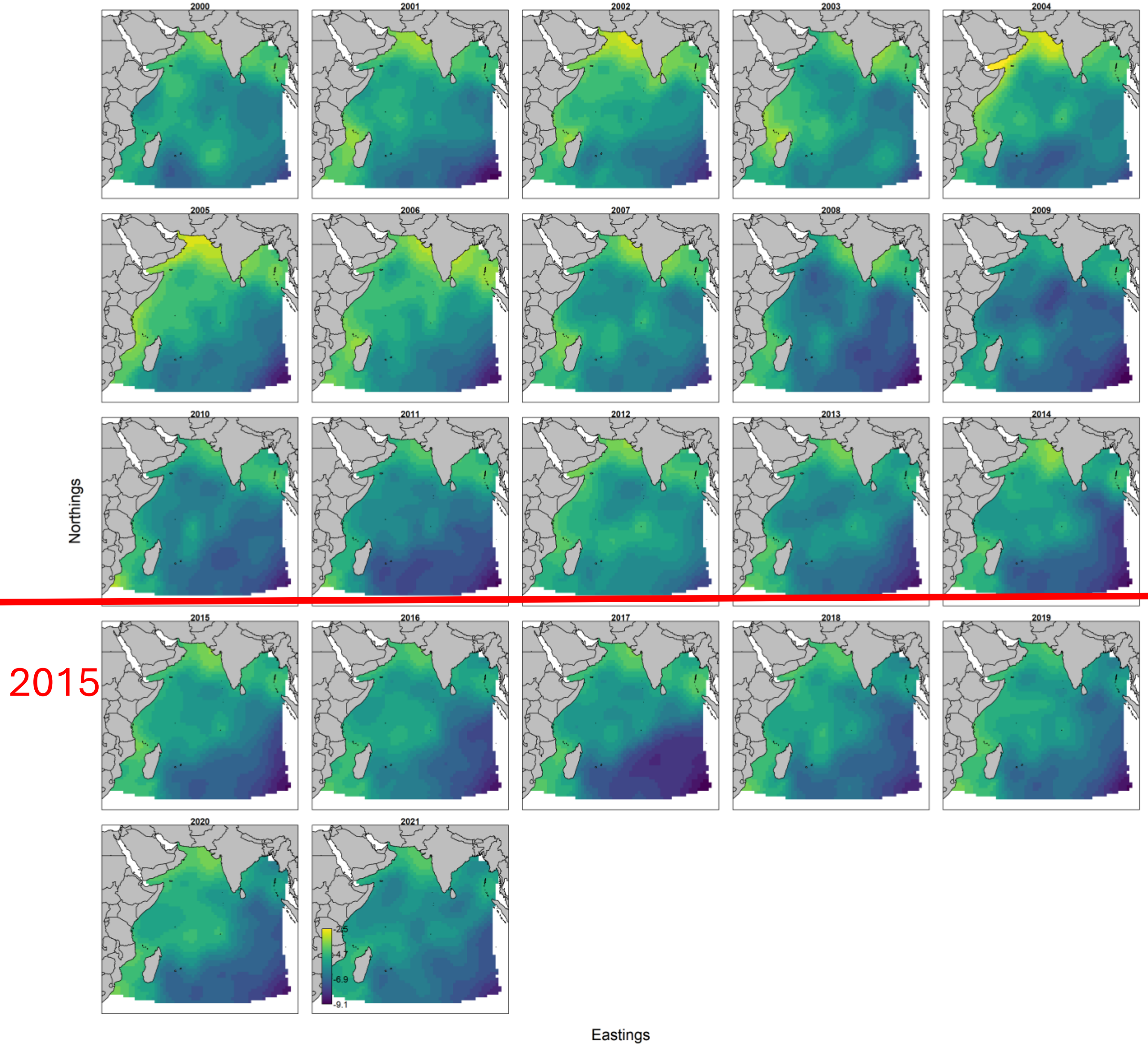


Base_exST
(exclude SpatialTemporal random effects)

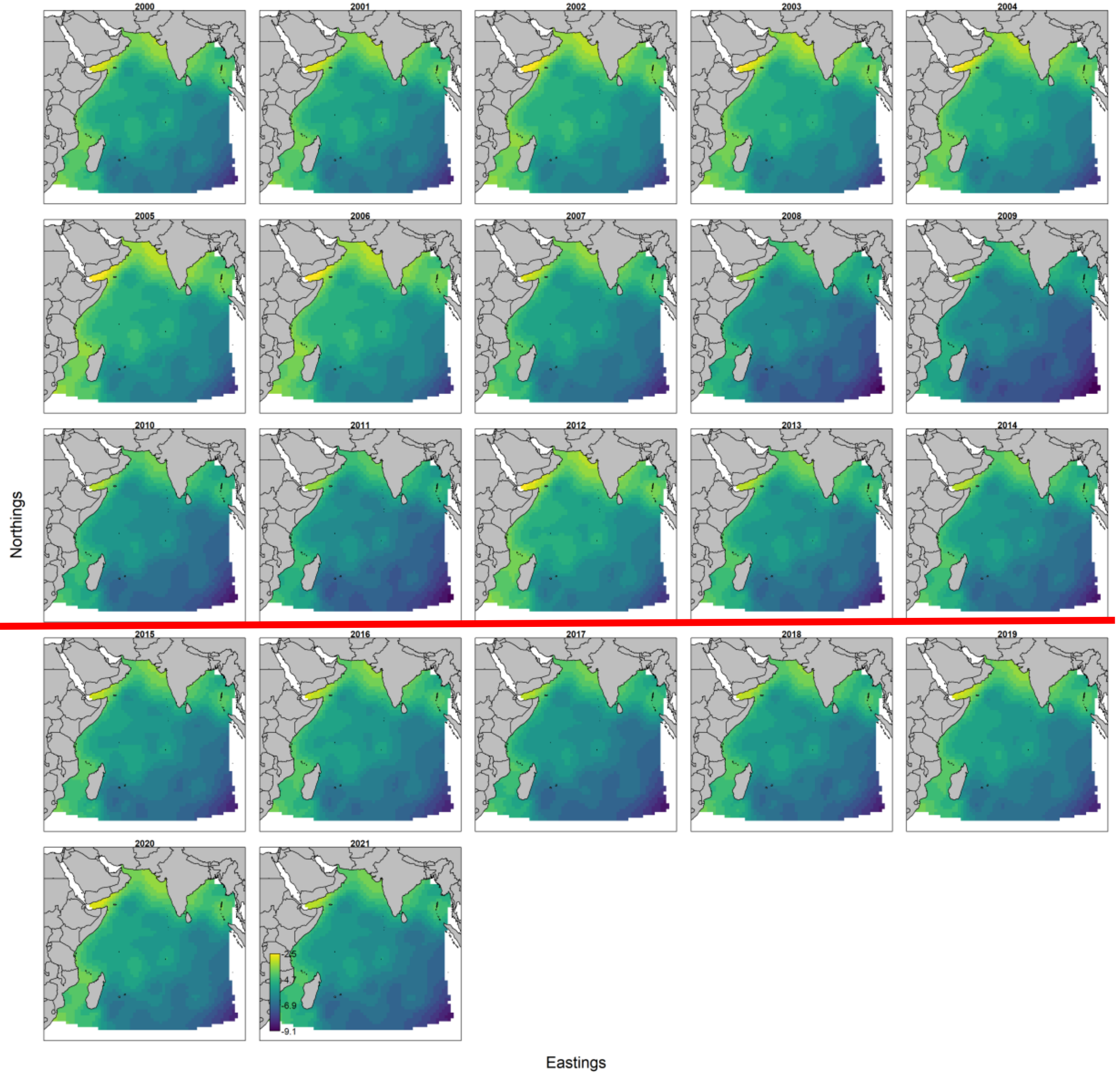


Large differences in indices (especially from 2010) attributable to spatial-temporal variation in abundance.

Base model (Season 1)
(include SpatialTemporal random effects)

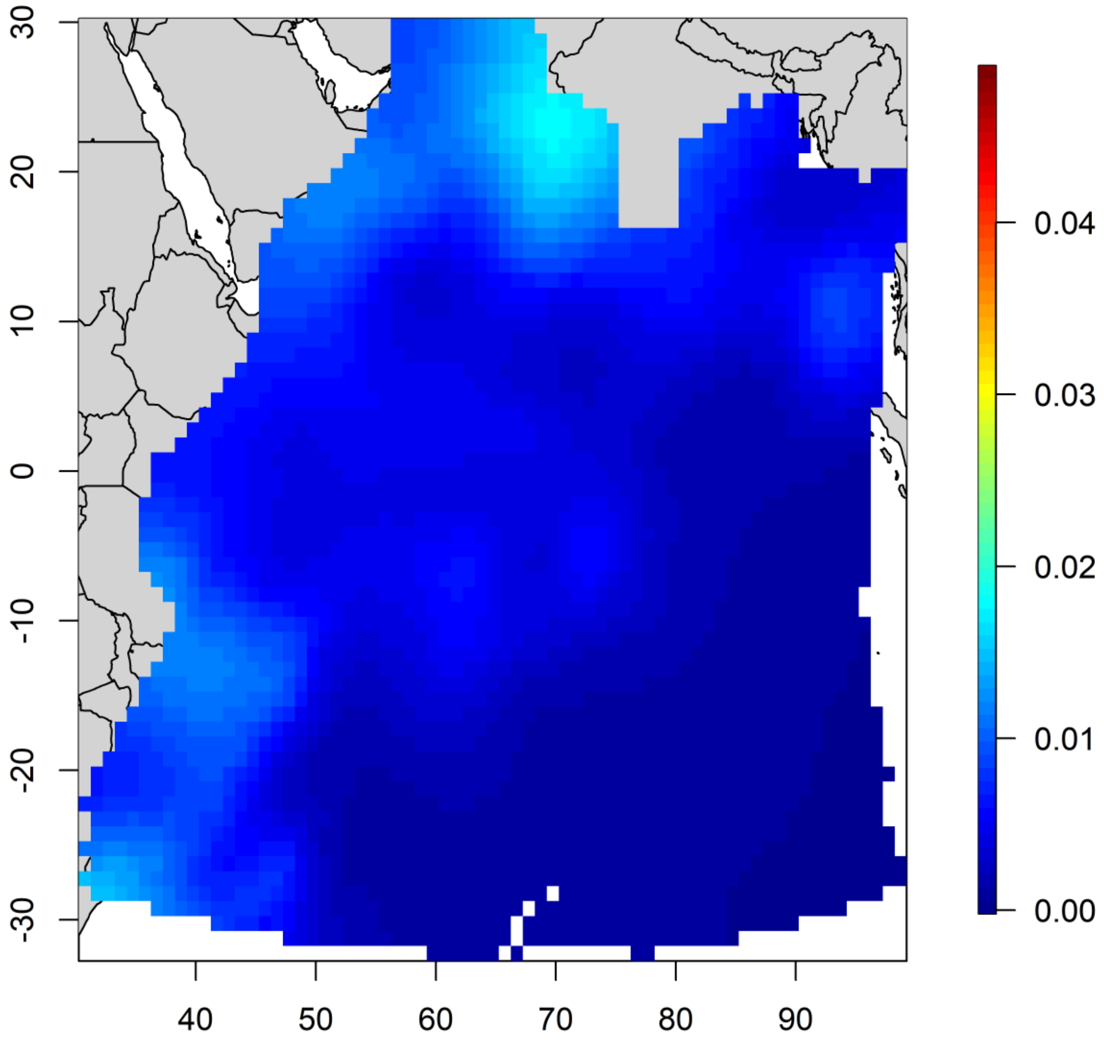


Base_exST
(exclude SpatialTemporal random effects)

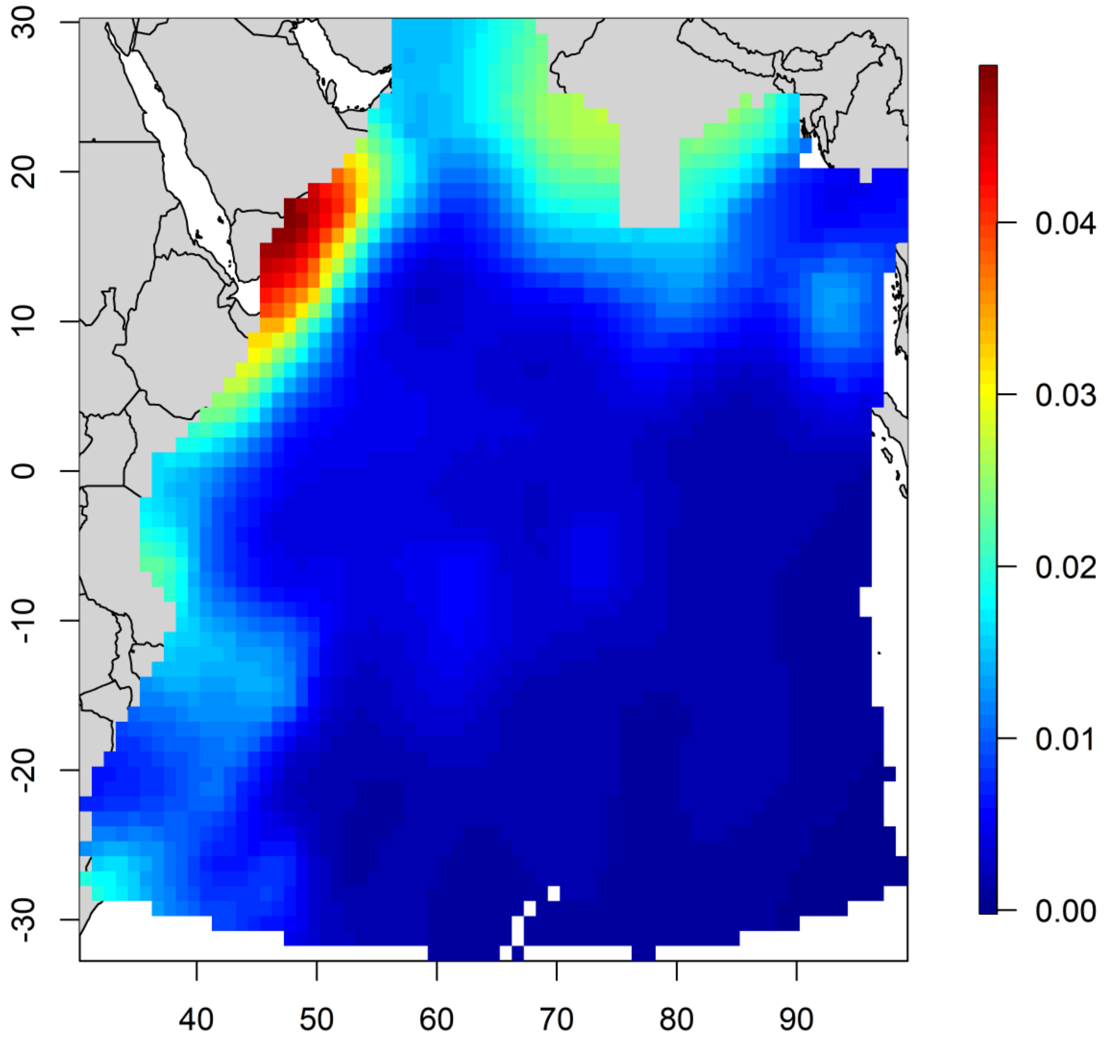


- Where are the differences in the spatial effects between the two models?
- Average density 2015-2021

Base model (Season 1)
(include SpatialTemporal random effects)



Base_exST
(exclude SpatialTemporal random effects)



Where are the differences in the spatial effects between the two models?

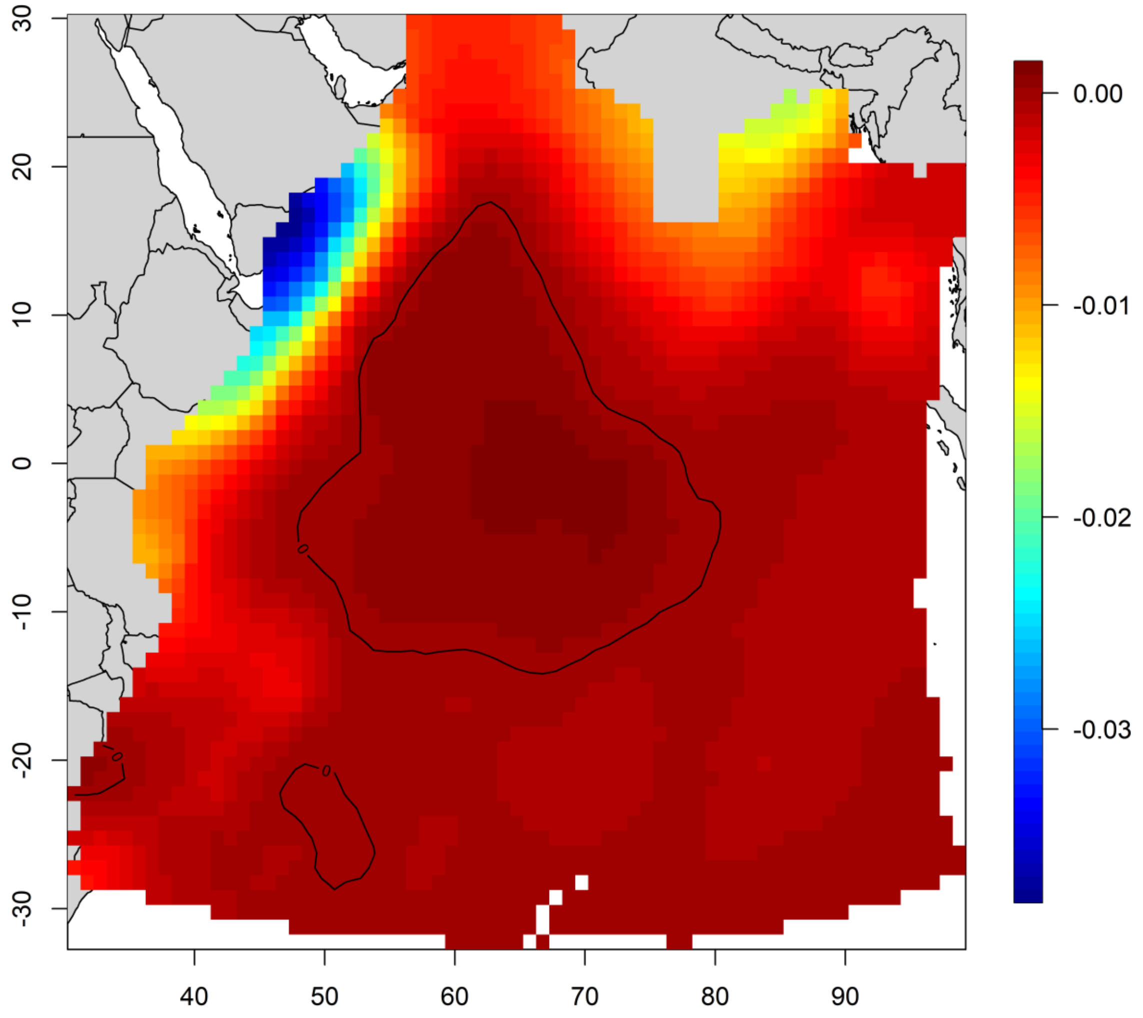
Differences in predicted density from the two models, aggregated 2015-2021 (Base – Base_exST).

Contour line = 0.0 i.e. no difference between models.

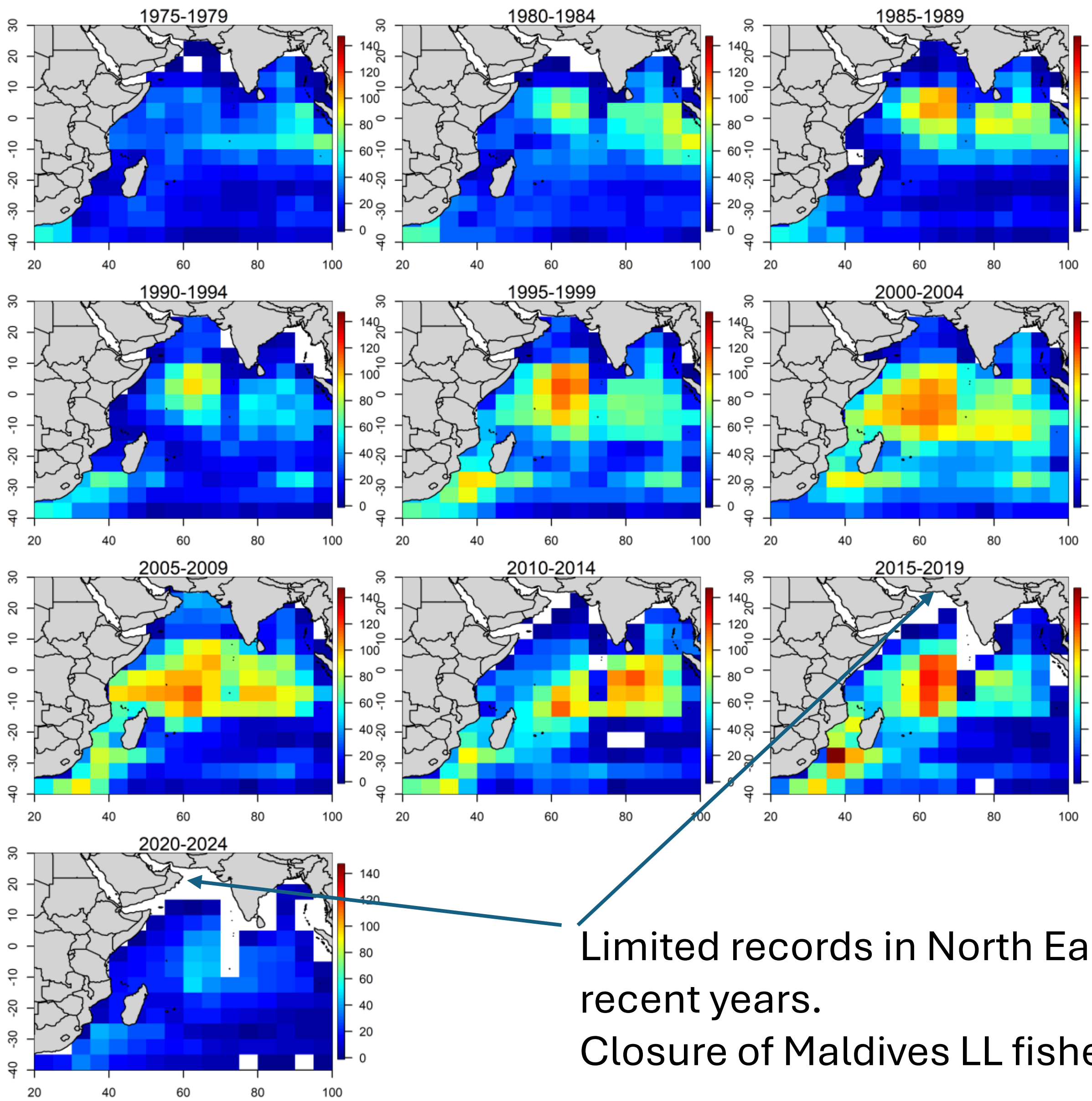
Overall difference between the two models appears related to the area where there has been a reduction in data. Informed by earlier years. Main differences off the coasts of Somali, Yemen and Oman. and Bay of Bengal.

Base model is estimating low spatial effects for those years.

(Equivalent result when include environmental covariates).



Number of Records



Limited records in North East IO in recent years.
Closure of Maldives LL fishery in 2019.

Environmental covariates.

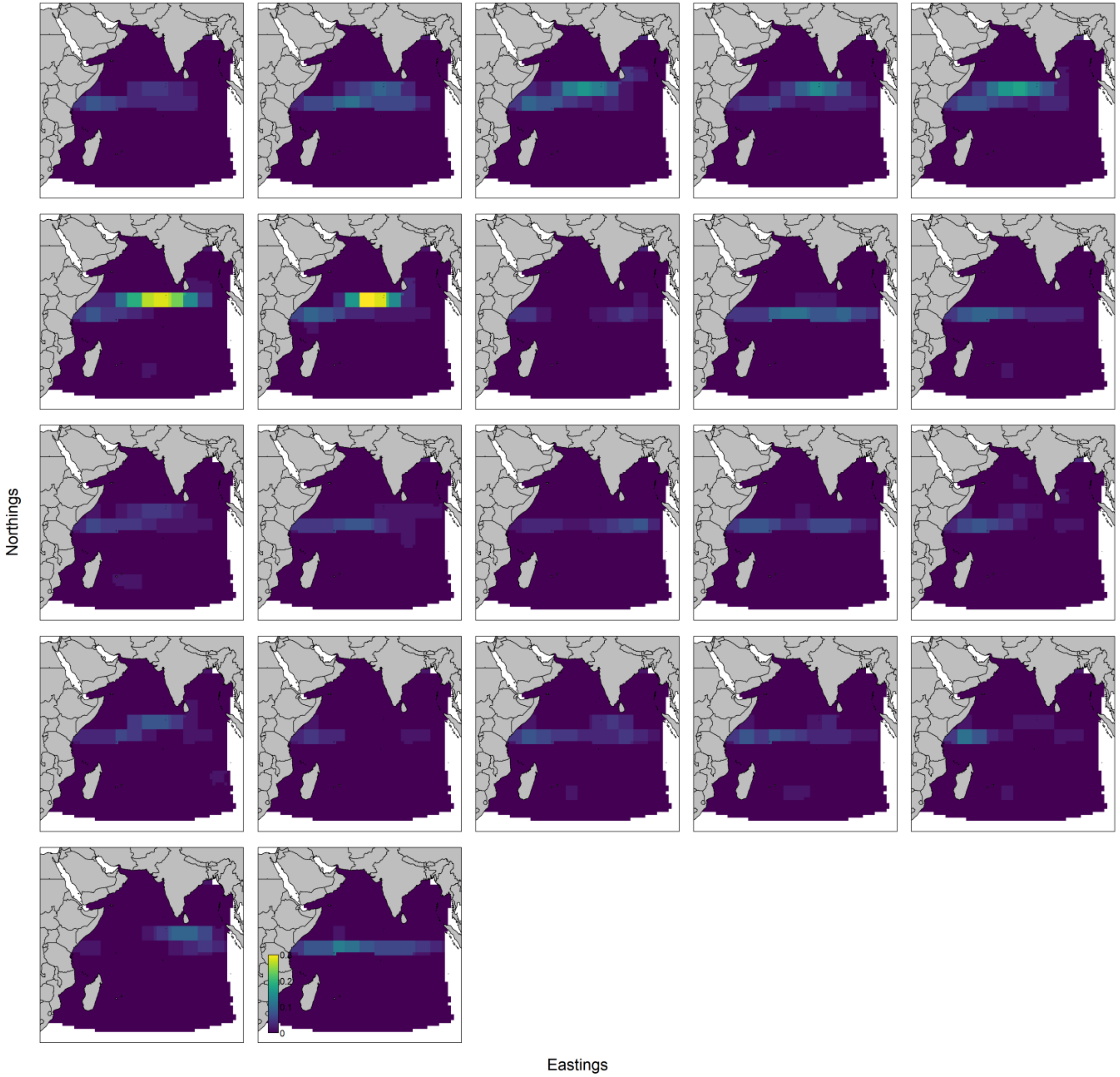
CurrentU (E-W), Mixed Layer Depth, Sea Surface Temperature.

Derived from NOAA NCEP EMC CMB GLOBAL Reyn_SmithOIv2 model.

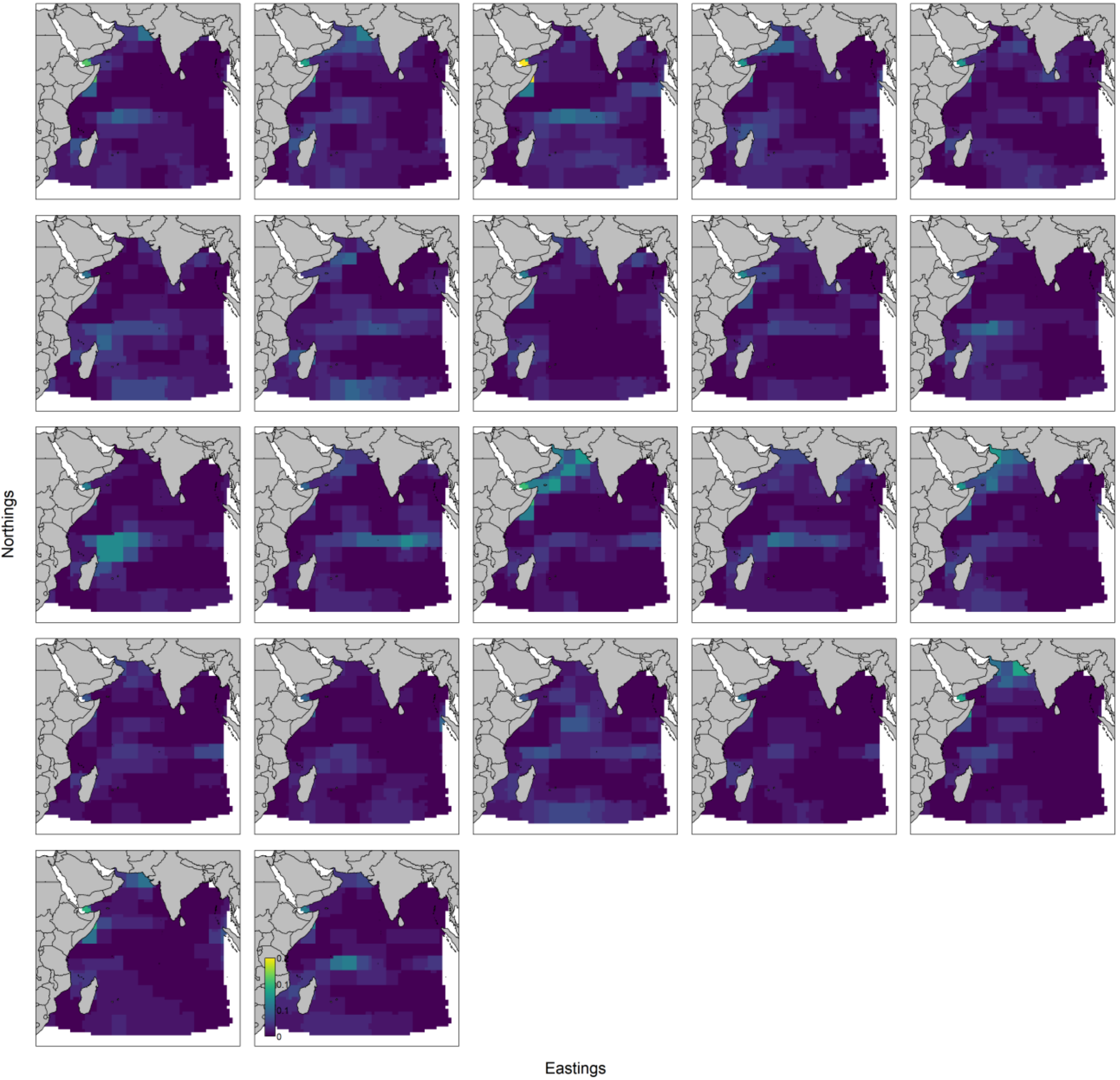
Data aggregated by year, season, Lat5, Long5.

Reynolds, R.W., N.A. Rayner, T.M. Smith, D.C. Stokes, and W. Wang, 2002: An Improved In Situ and Satellite SST Analysis for Climate. J. Climate, 15, 1609-1625.
Behringer, D. W., and Y. Xue, 2004: Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean. Eighth Symposium on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface, AMS 84th Annual Meeting, Washington State Convention and Trade Center, Seattle, Washington, 11-15.

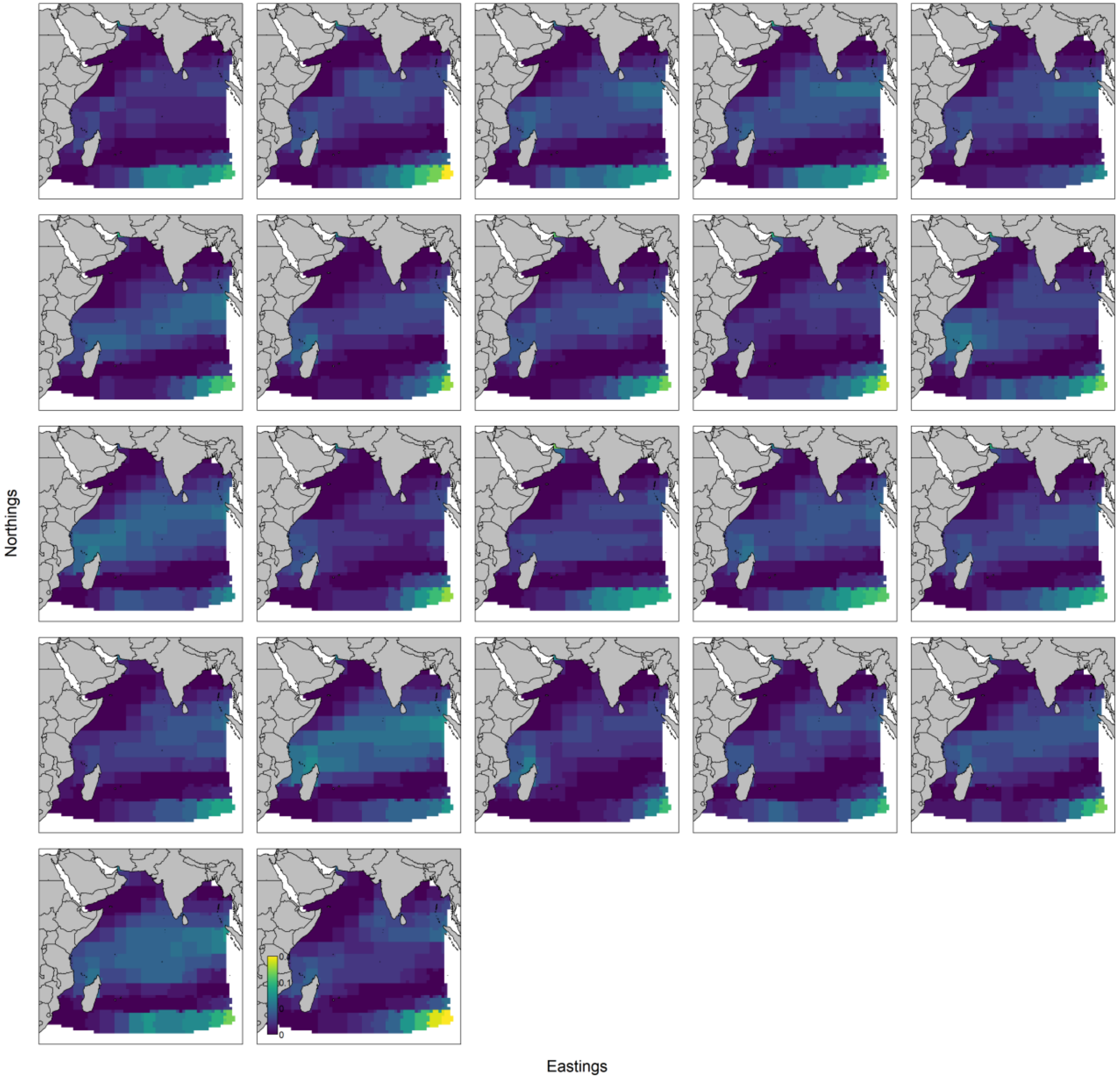
CurrentU covariate



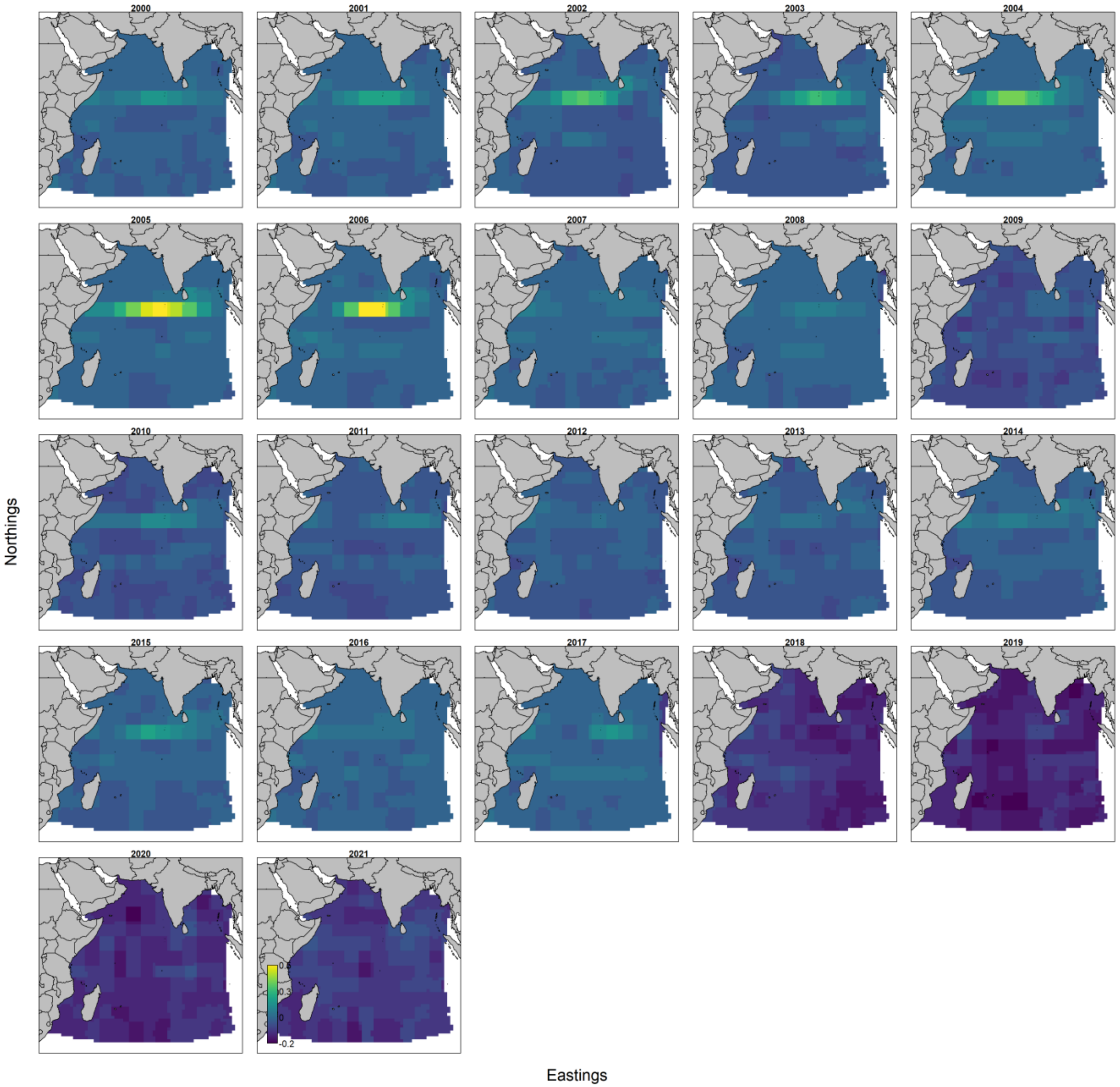
MLD covariate



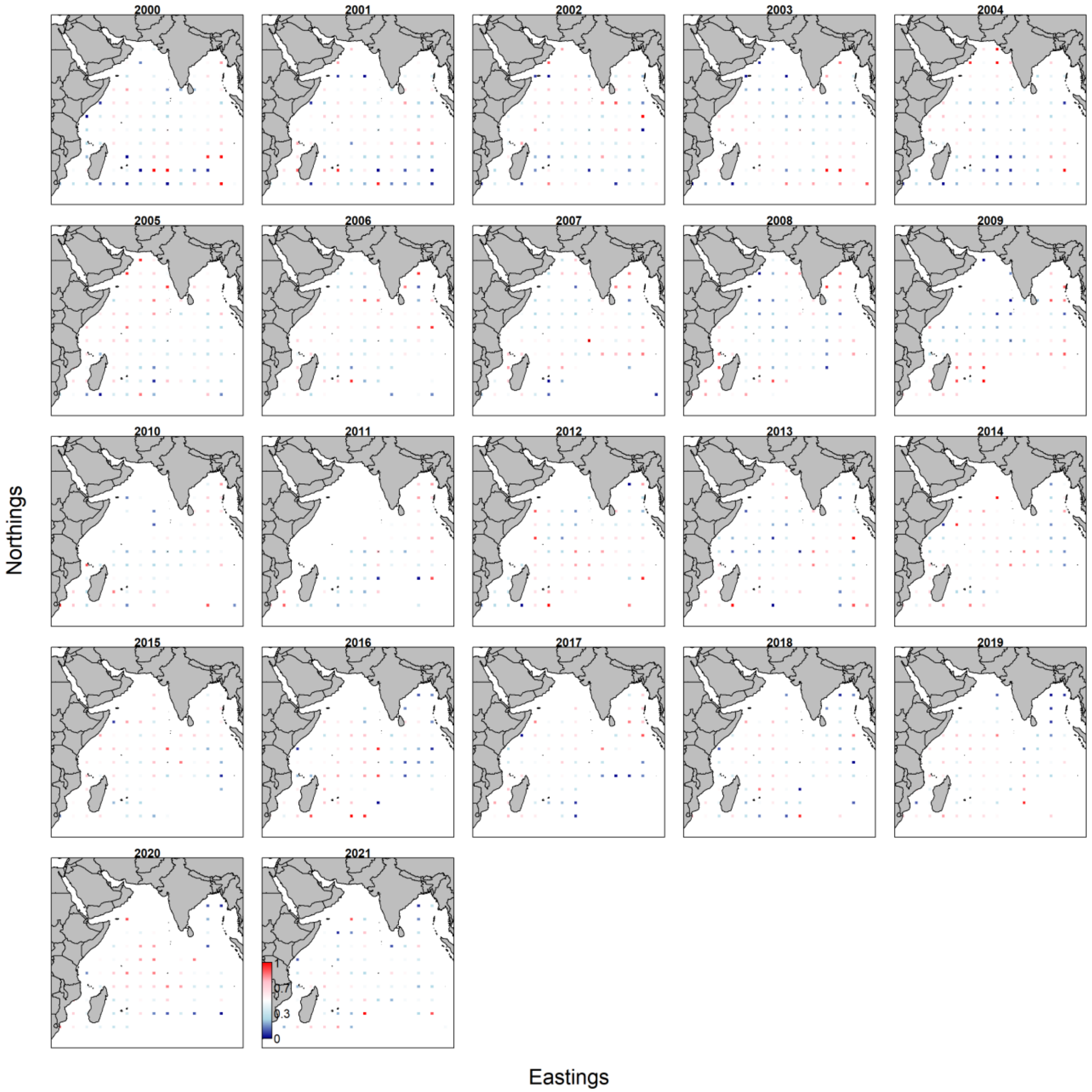
SST covariate



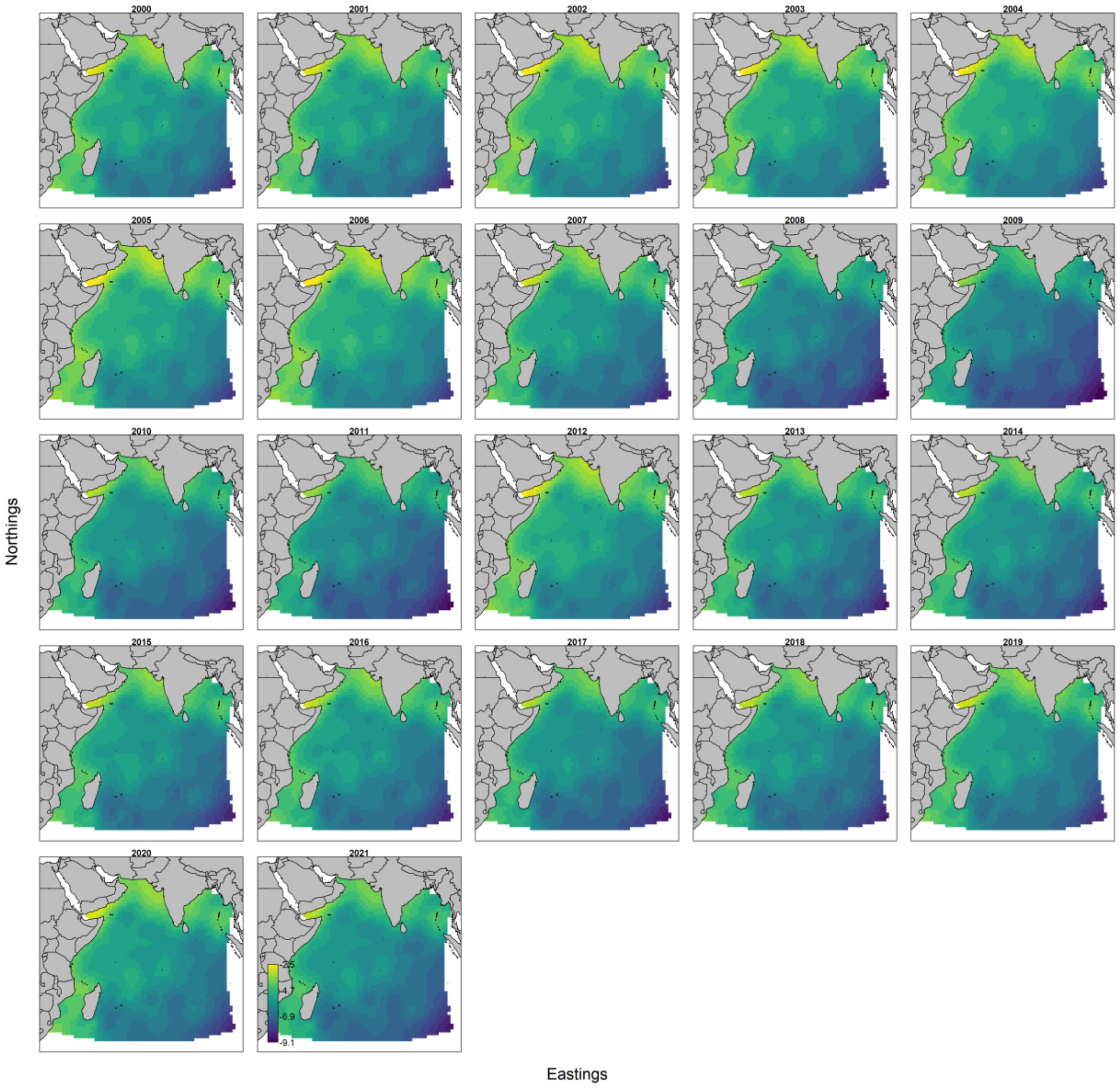
Base_exST + CurrentU + SST + MLD
(exclude SpatialTemporal random effects)
Combined covariate effects



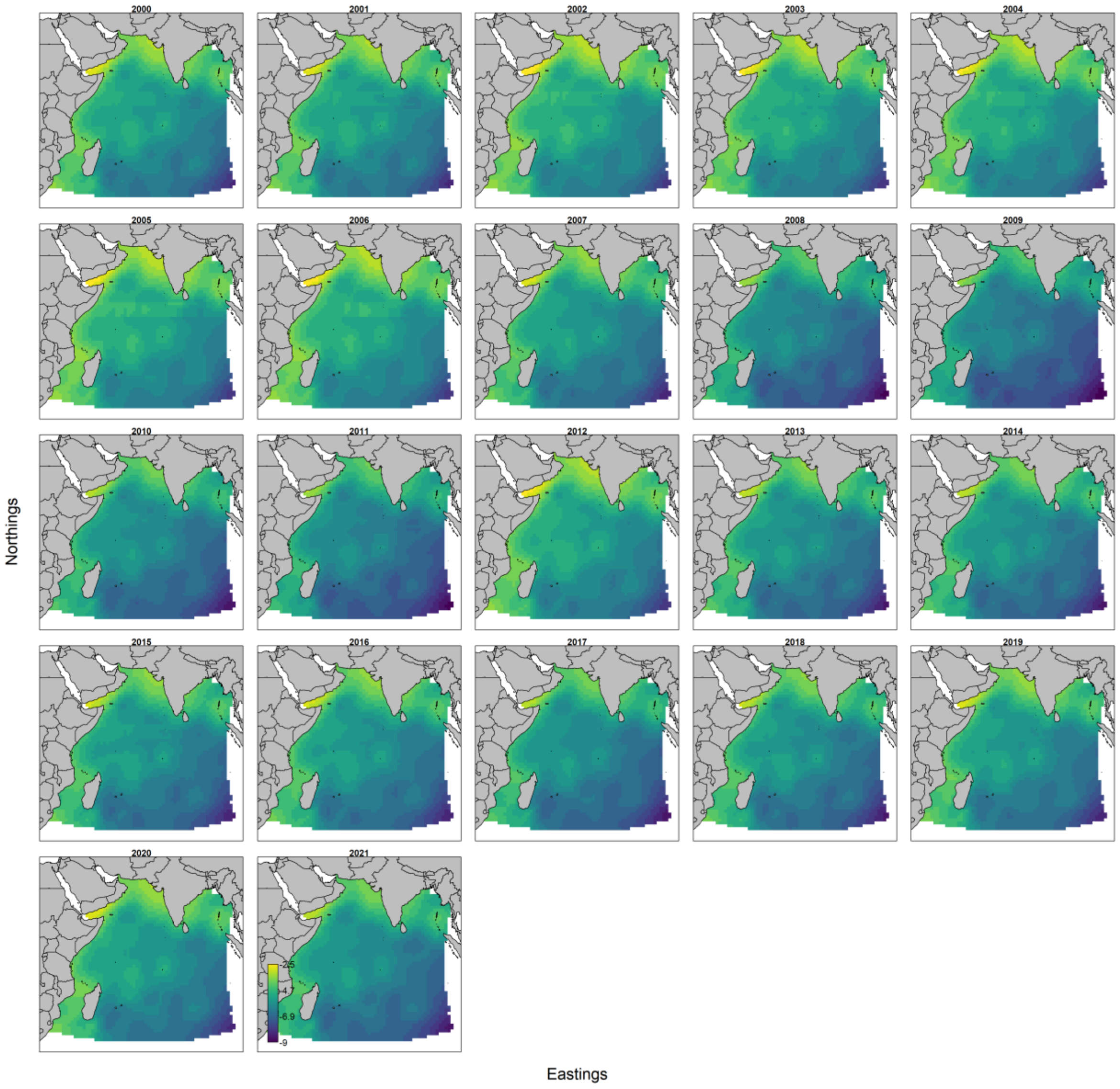
Base_exST + CurrentU + SST + MLD
(exclude SpatialTemporal random effects)
Residuals

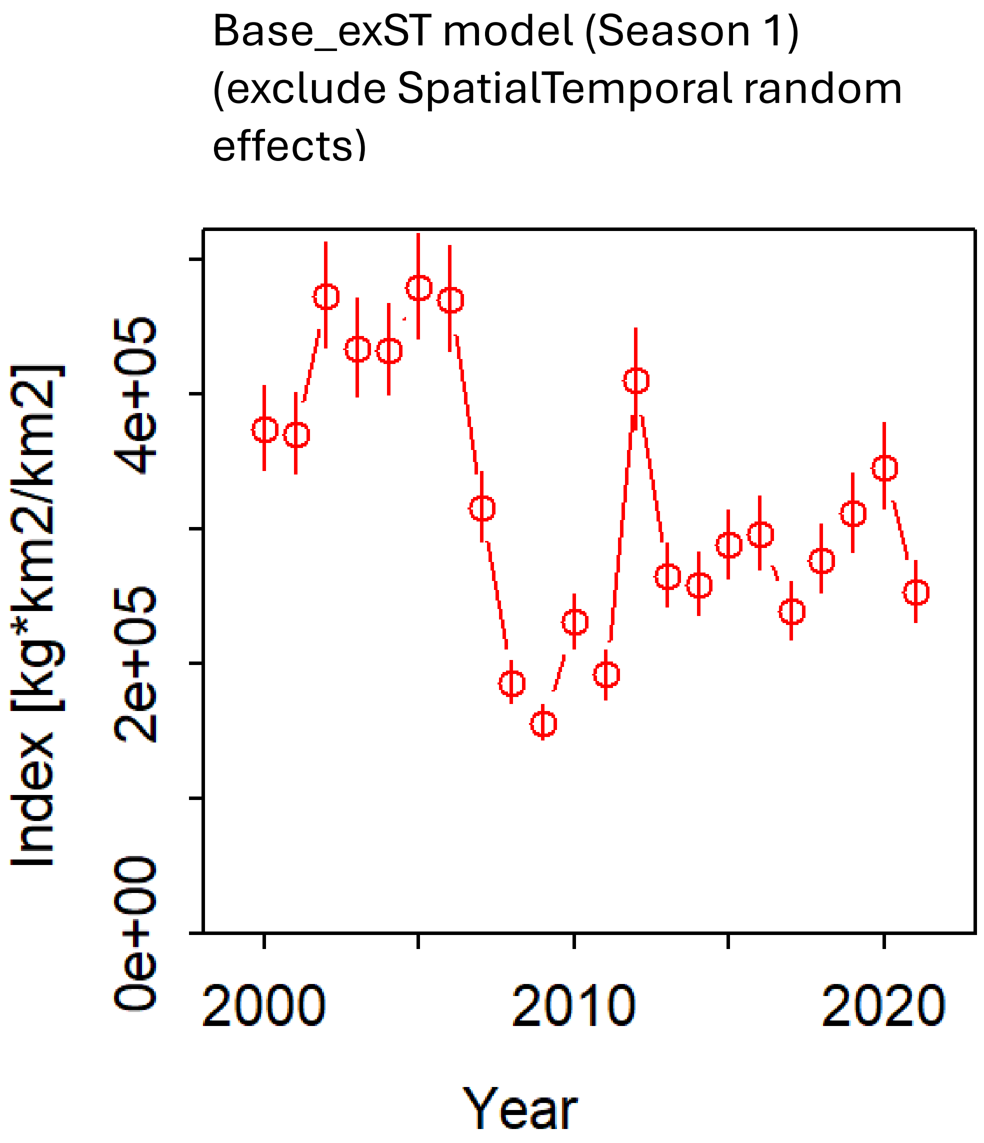


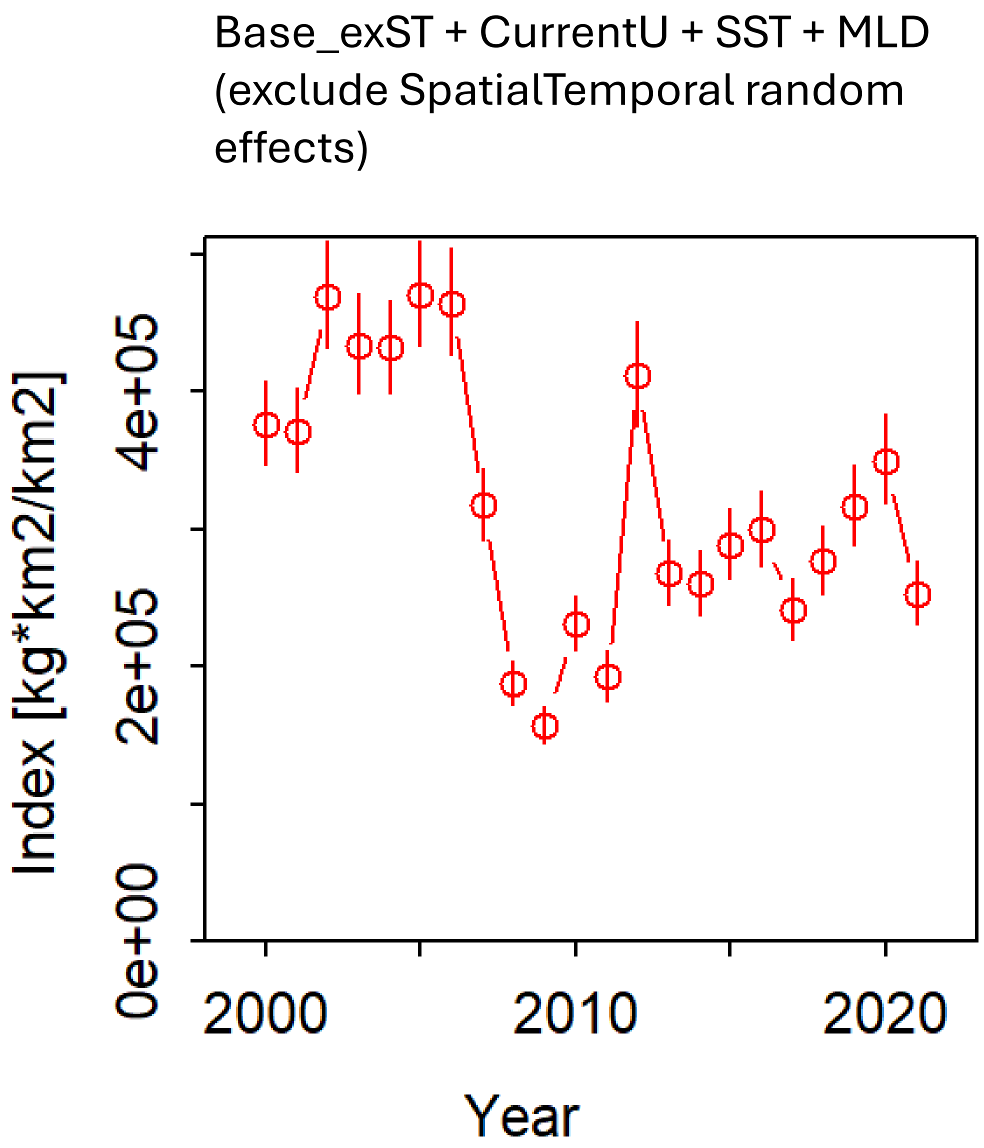
Base_exST
(exclude SpatialTemporal random effects)
Predicted density



Base_exST + CurrentU + SST + MLD
(exclude SpatialTemporal random effects)
Predicted density







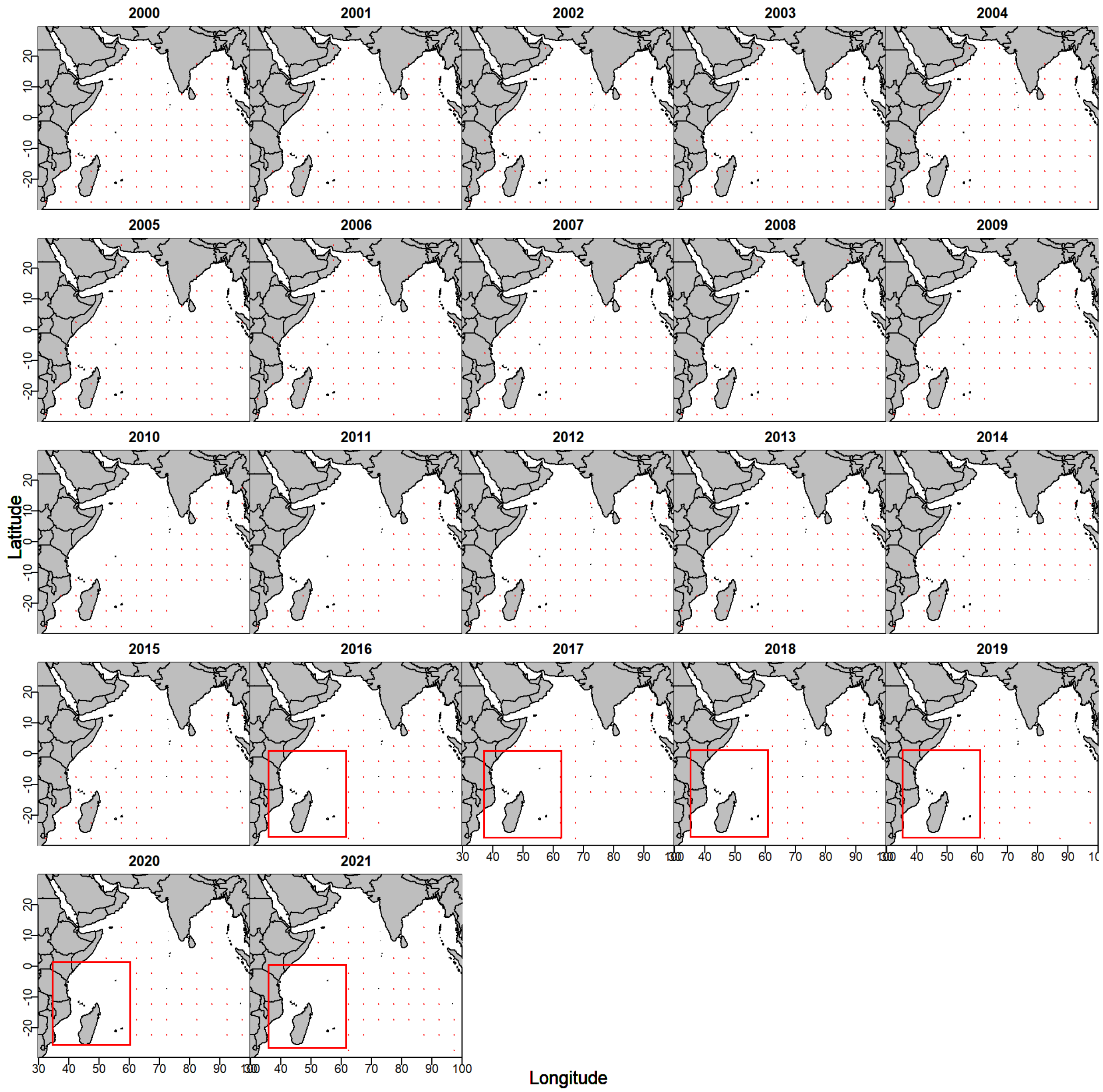
No appreciable differences in annual indices with inclusion of environmental covariates.
Environmental covariates <u>*not*</u> *accounting for spatial-temporal variation in CPUE.*

Model testing

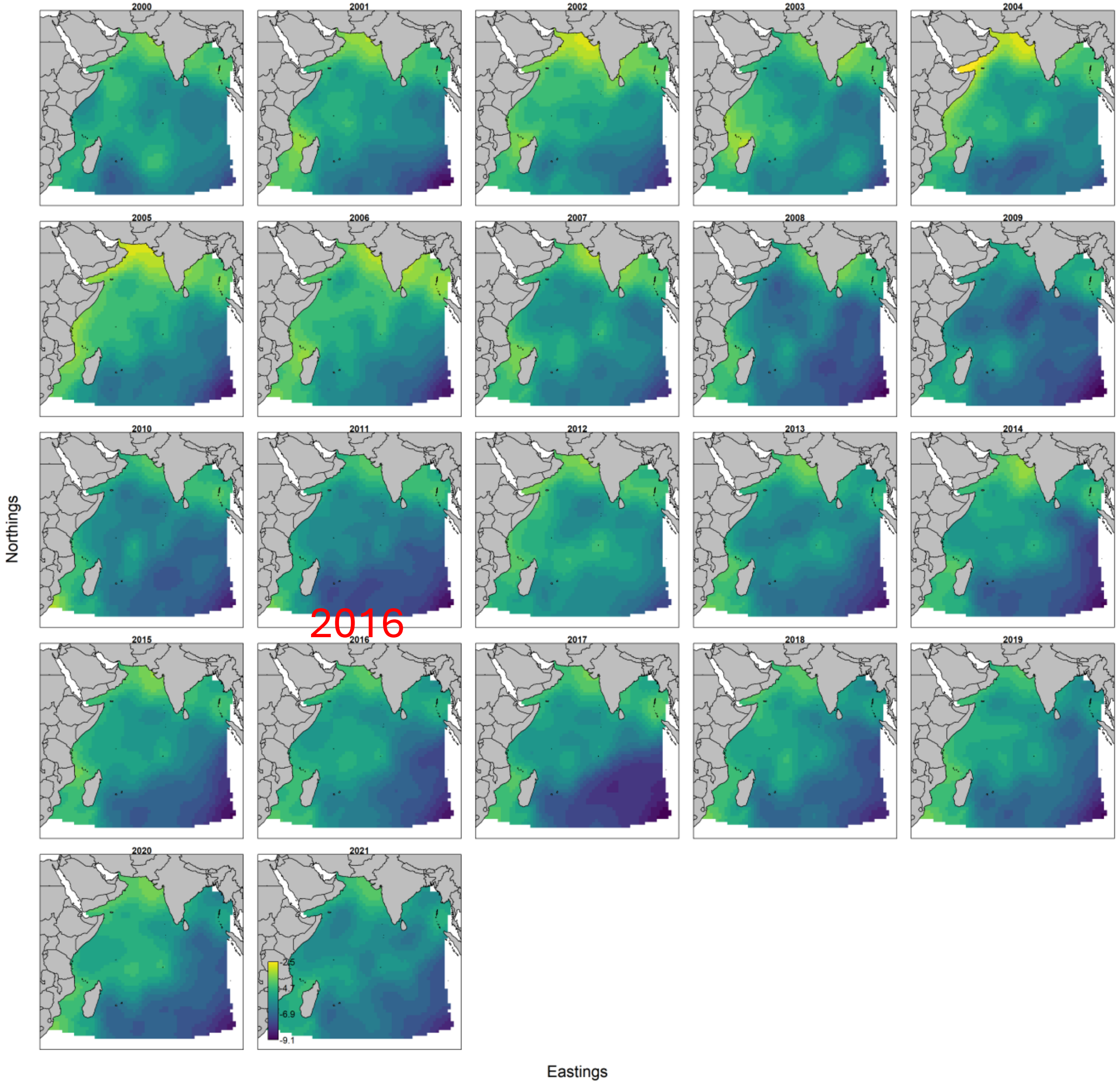
Evaluate impact of “missing data”.

Repeat Base model, Season 1 excluding data from SE Indian Ocean from 2016-2021.

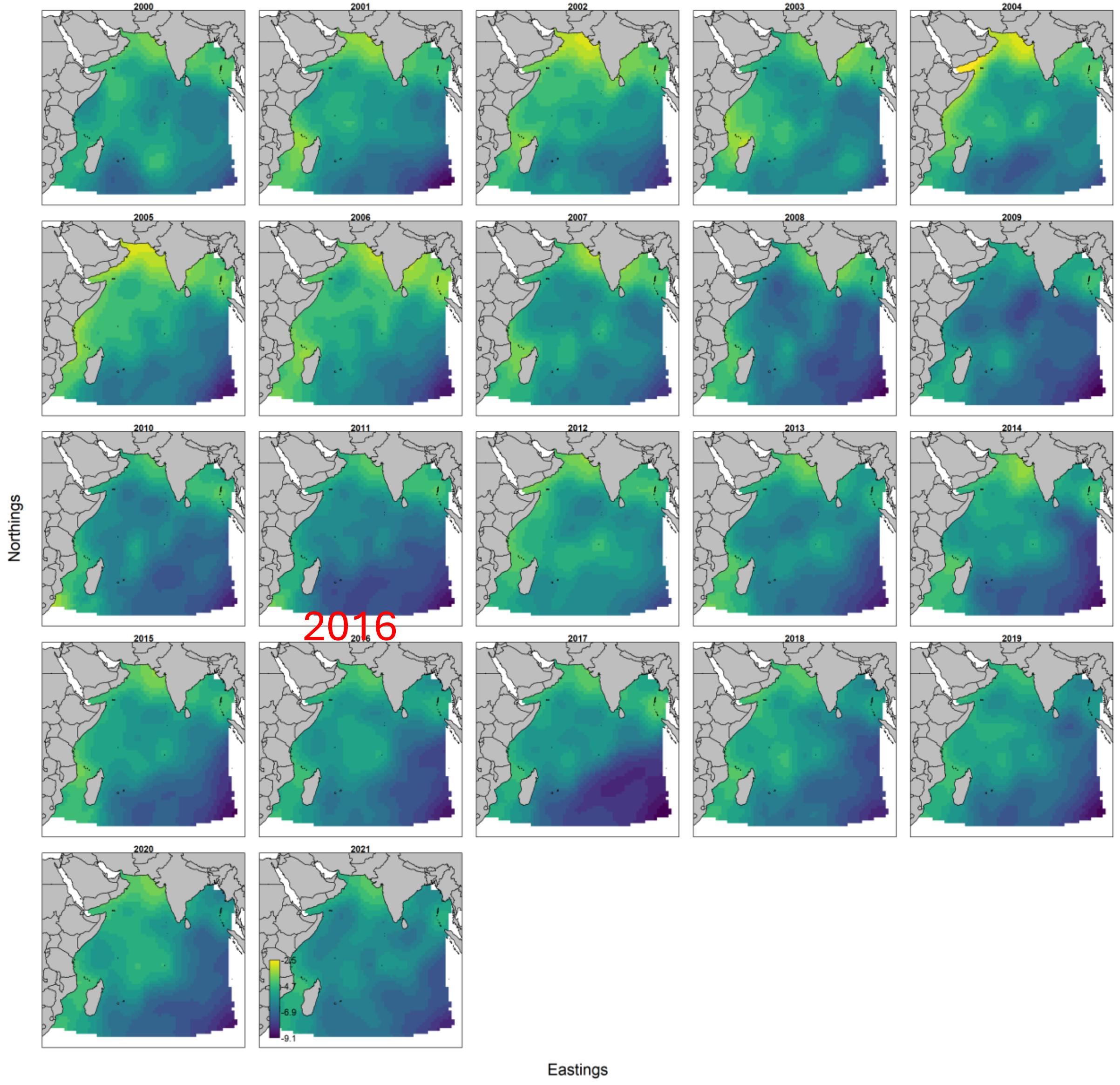
Compare resulting estimates of spatial density and overall IO indices.



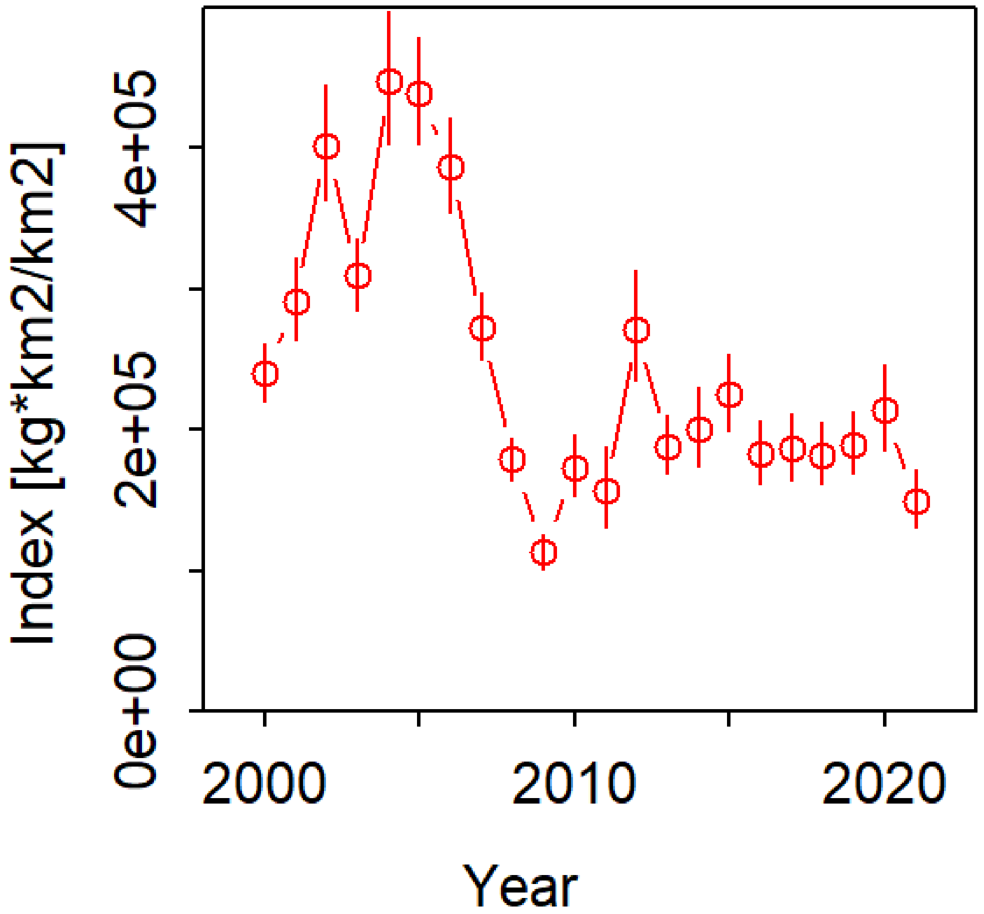
Base model (Season 1)
(include SpatialTemporal random effects)



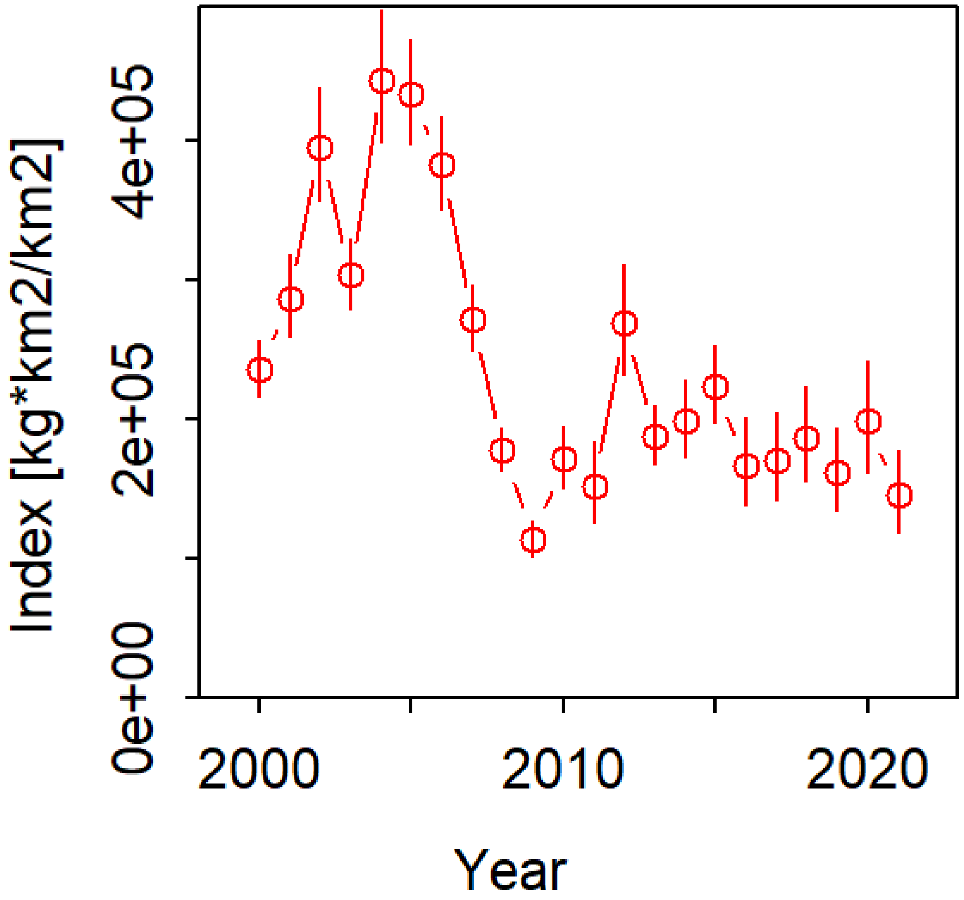
Test
(include SpatialTemporal random effects)



Base model (Season 1)
(include SpatialTemporal random effects)

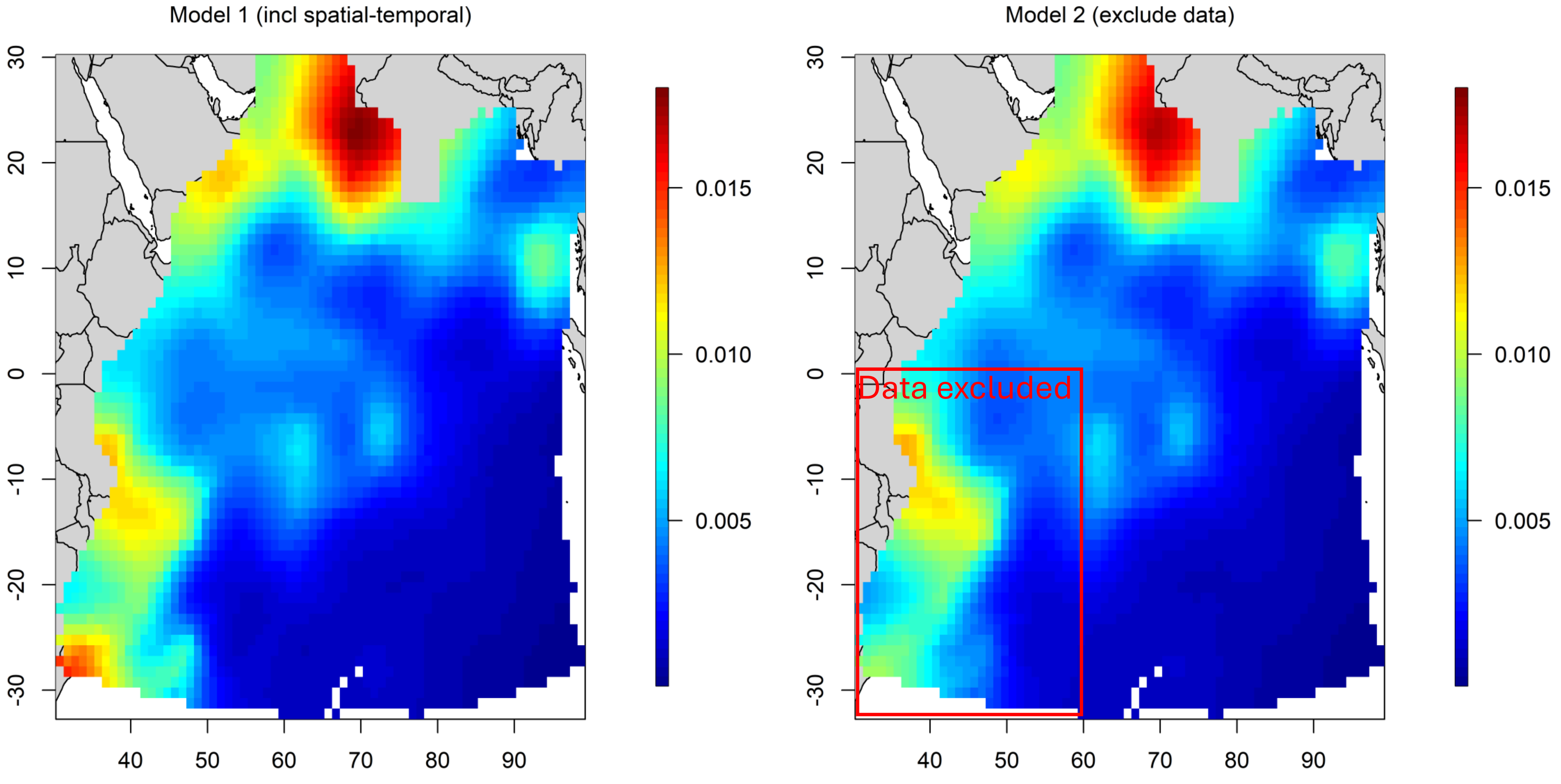


Test
(include SpatialTemporal random effects)



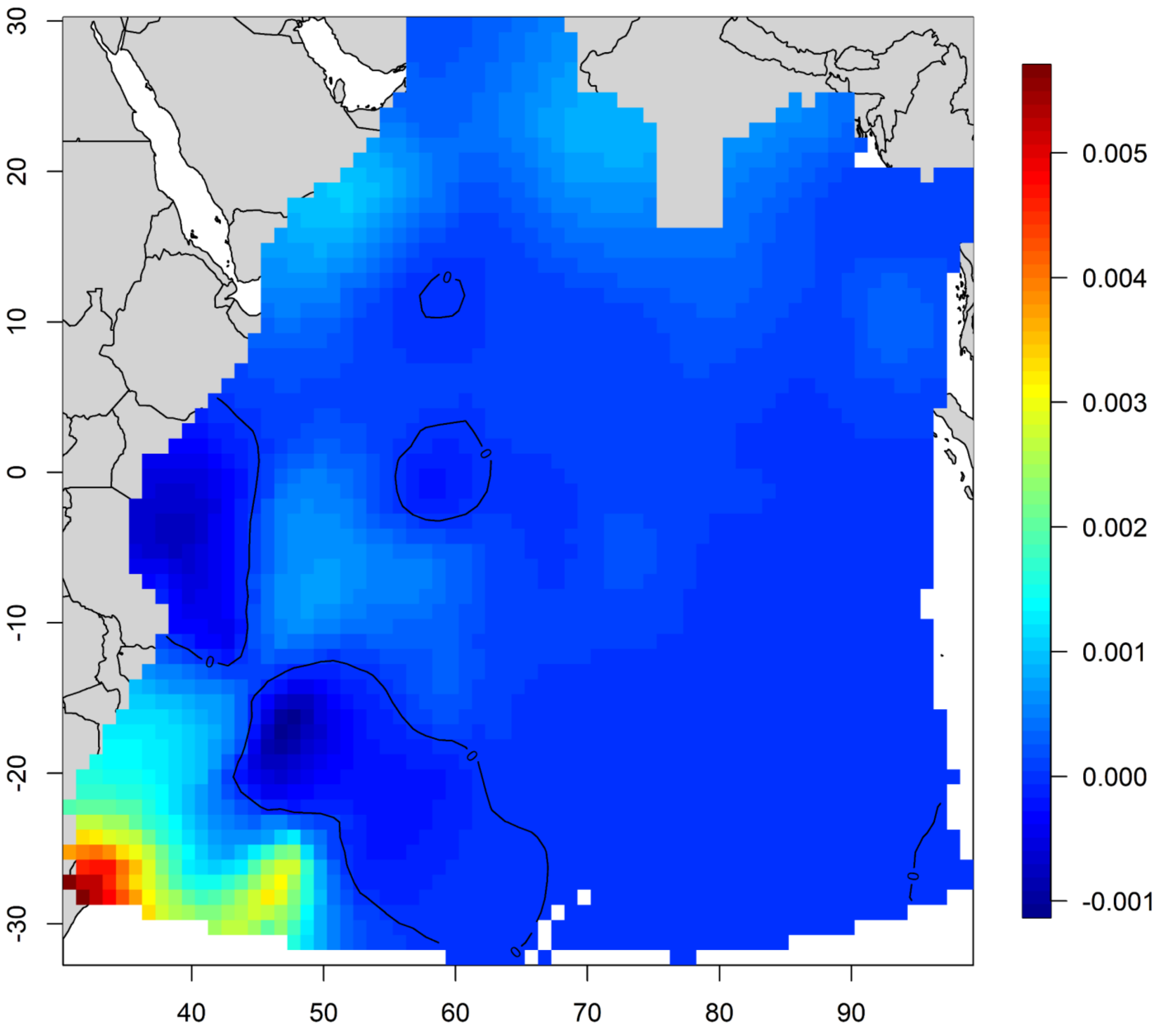
Indices are equivalent from 2000-2015.
Indices from 2016-2021 are about 10% lower when data excluded (right figure).
Higher uncertainty in 2016-2021 indices with data excluded (right figure).

Comparison of average density from 2016-2021 for full data set (left) and with data excluded (right).



Estimates of density are very similar between the two models.

Only appreciable difference is in the extreme SE corner – high density predicted by full data set.


30
20
10
0
-10
-20
-30
40
50
60
70
80
90
0.005
0.004
0.003
0.002
0.001
0.000
-0.001

Summary 2

- Reliability of VAST predictions when extrapolating from adjacent observations? Initial trial result was encouraging. Further evaluation via model testing.
- Environmental covariates may be more informative when included as environmental drivers (appropriate time lag), but difficult to implement within VAST framework.
- Consideration of seasonal, spatial interaction in GLM.