Multivariate time-series forecasting of the NE Arabian Sea Oil Sardine fishery using satellite covariates

Eli Holmes[1], Nimit Kumar[2], Sourav Maity[2], B.R Smitha[3], Sherine Cubelio[3], Cara Wilson[4], Narayanane Saravarane[2], Nagaraja Kumar[3]

1. NW Fisheries Science Center, NOAA Fisheries, Seattle, WA USA
2. Indian National Centre for Ocean Information Services, Hyderabad, India
3. Centre for Marine Living Resources & Ecology, Kochi, India
4. SW Fisheries Science Center, NOAA Fisheries, La Jolla, CA USA

faculty.washington.edu/eeholmes

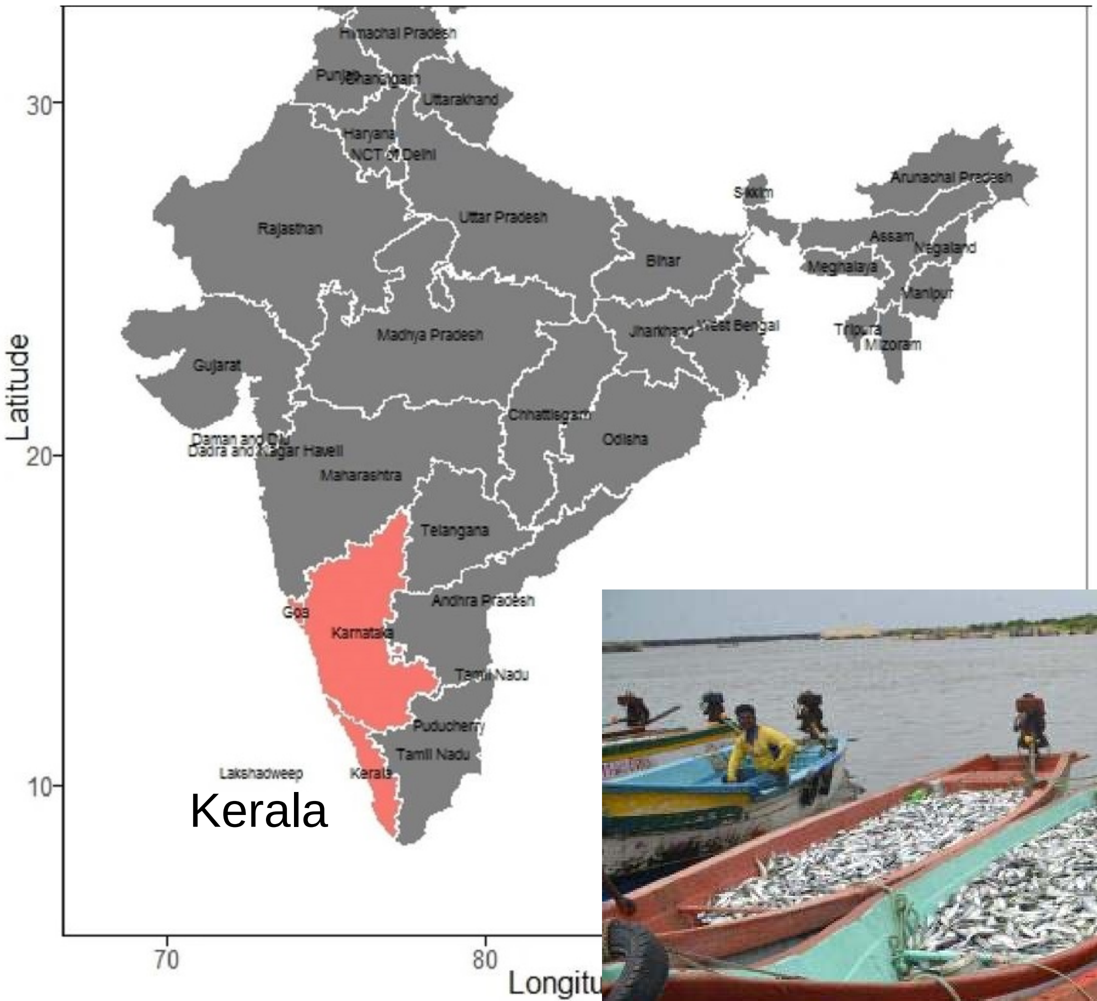
MoES-NOAA Collaboration: Development of Predictive Capabilities for Marine Fisheries and Harmful Algal Blooms in Indian Seas

- **Research collaboration on forecasting**:
  - Harmful Algal Blooms
  - ***Oil sardine fishery***

INCOIS: Indian National Centre for Ocean Information Services
http://www.incois.gov.in

CMFRE: Centre for Marine Living Resources & Ecology

Indian Oil Sardine Fishery

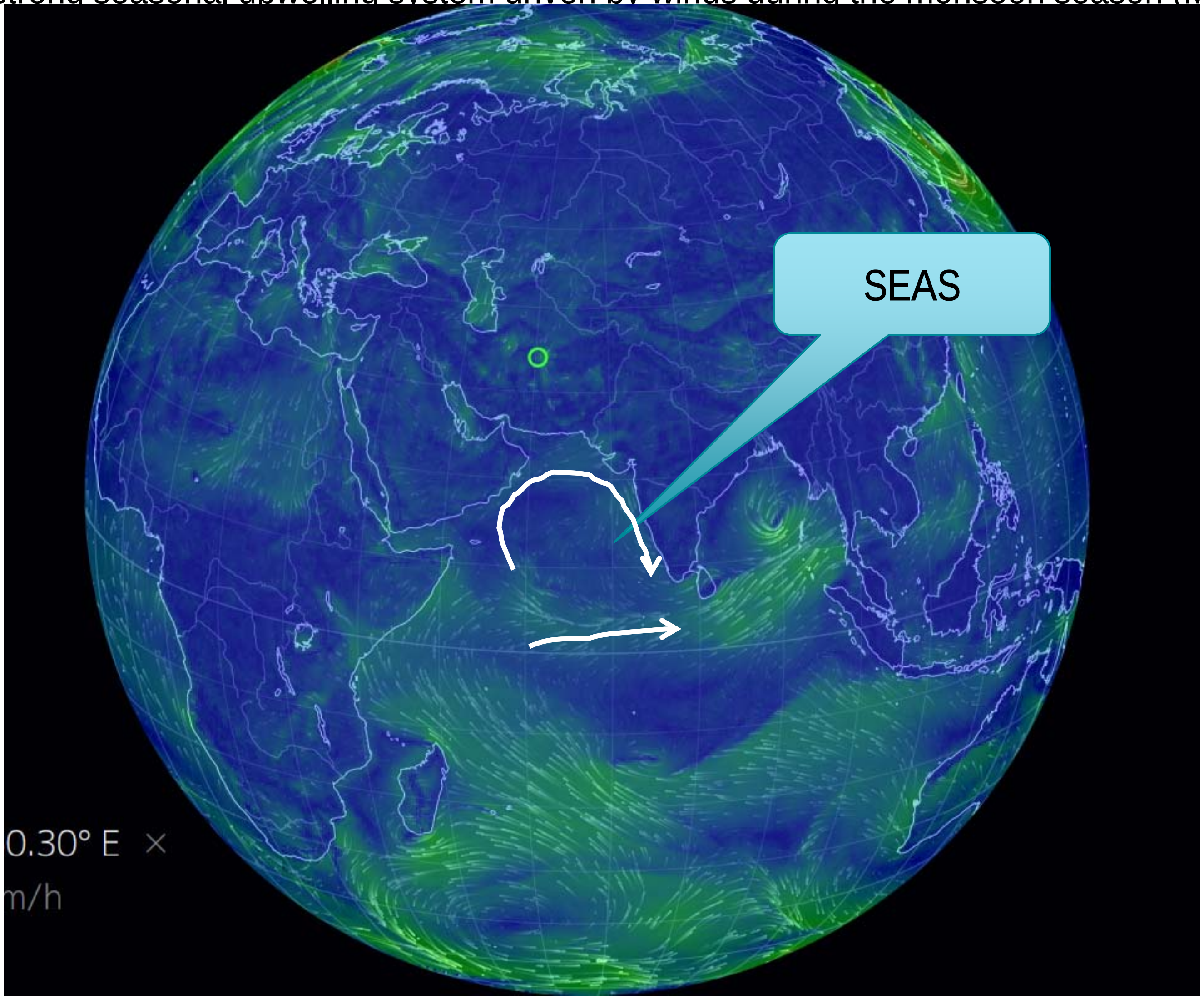


Today's talk --- overview of the project, challenges, and initial results

- Background
  - Physical processes (upwelling) affecting oil sardines in the SE Arabian Sea
  - Biology of oil sardines and how the interacts with the above
- Satellites covariates
- Exploratory correlation analysis
  - Which covariates have explanatory value? biology + physical processes + information in the covariate
- Testing some forecasting models
  - Regression models
  - Exponential smoothing models

The South East Arabian Sea is one of world's major upwelling zones and one of the most productive regions of the world's oceans

Has a strong seasonal upwelling system driven by winds during the monsoon season (May-Sept)



https://earth.nullschool.net/#current/wind/surface/level/orthographic=72.75,17.28,1041/loc=74.264,11.874

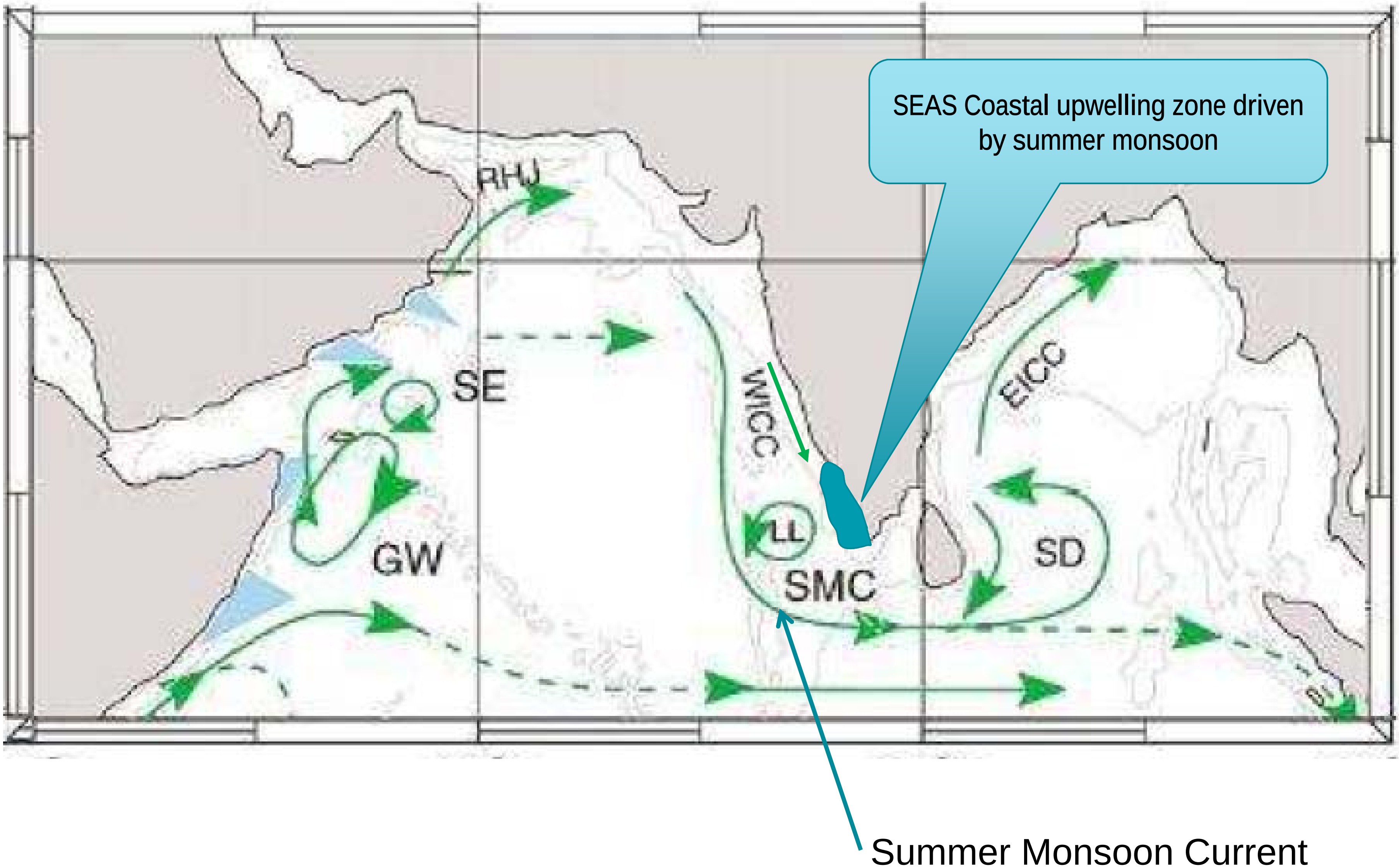
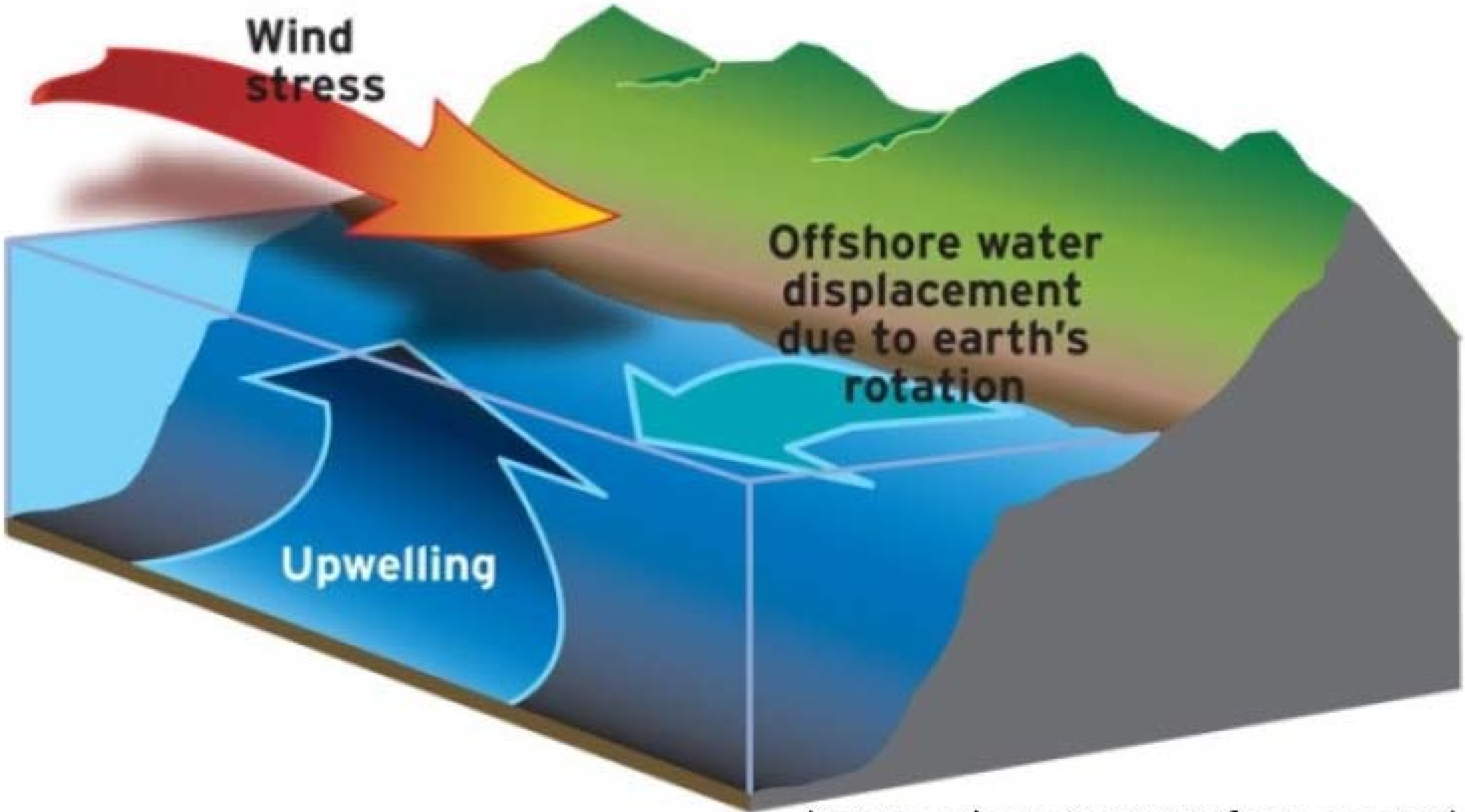


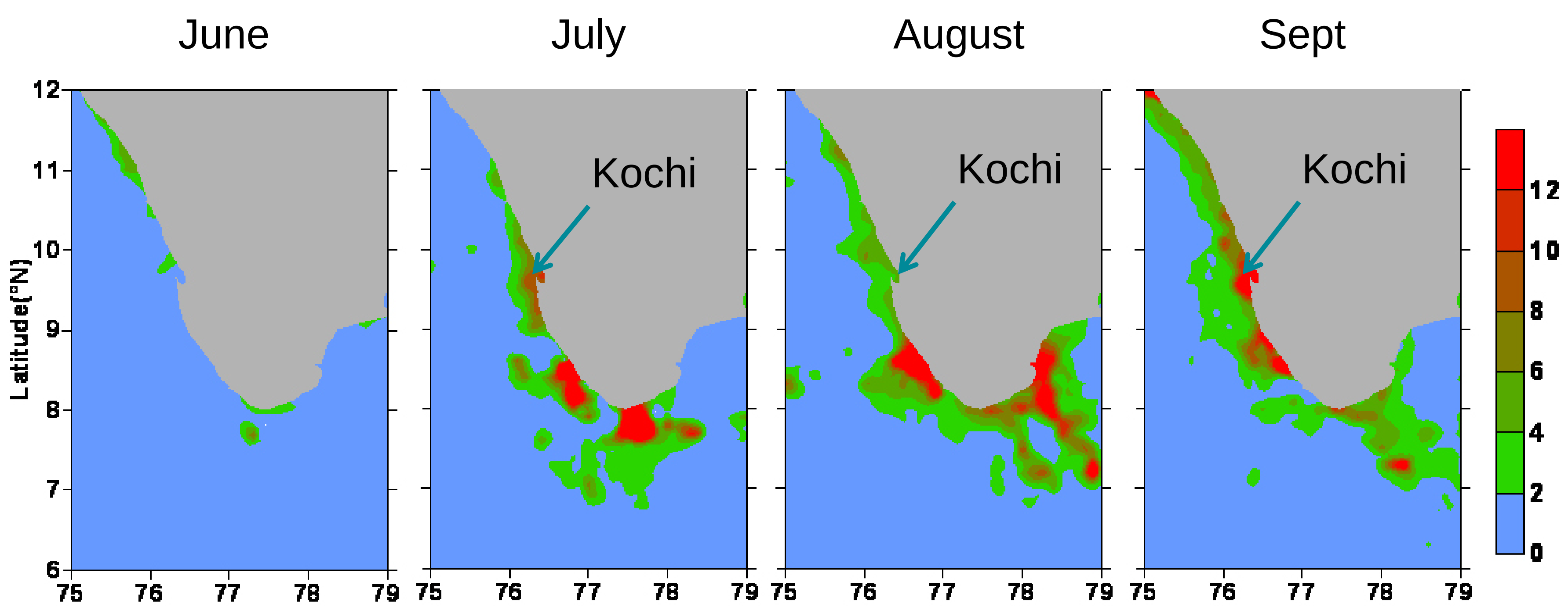
Figure 1.3 from Smitha BR, 2010, thesis



(Source : http://www.nwfsc.noaa.gov)

Our focus area: SW coast of India 8-13 deg latitude

Important oil sardine spawning area



Four-year (2003–2006) average monthly monthly surface Chlorophyll from MODIS AQUA for June–September.(monsoon months)

Figure 2.8 from Smitha BR, 2010, thesis

Life cycle of the oil sardine

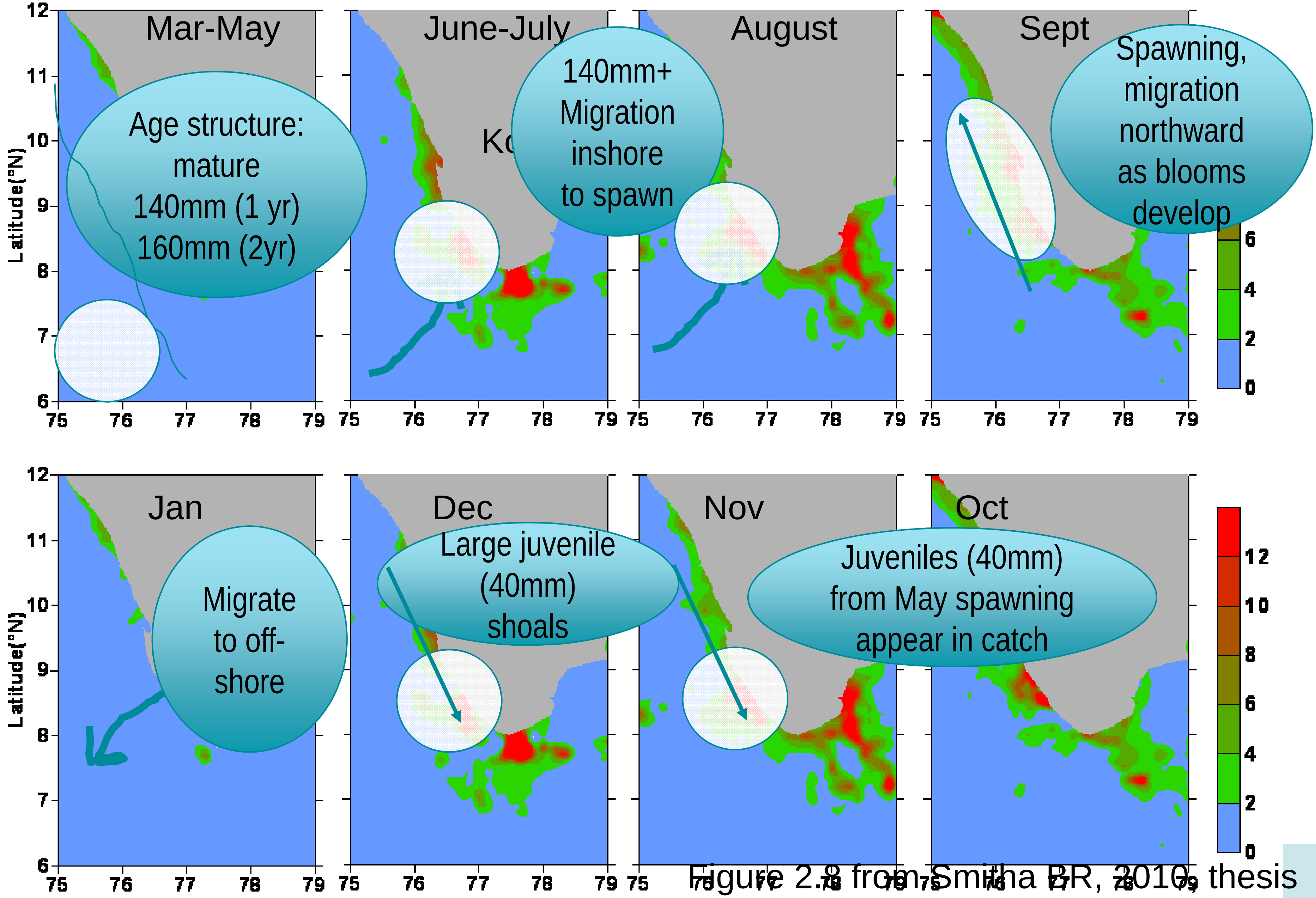
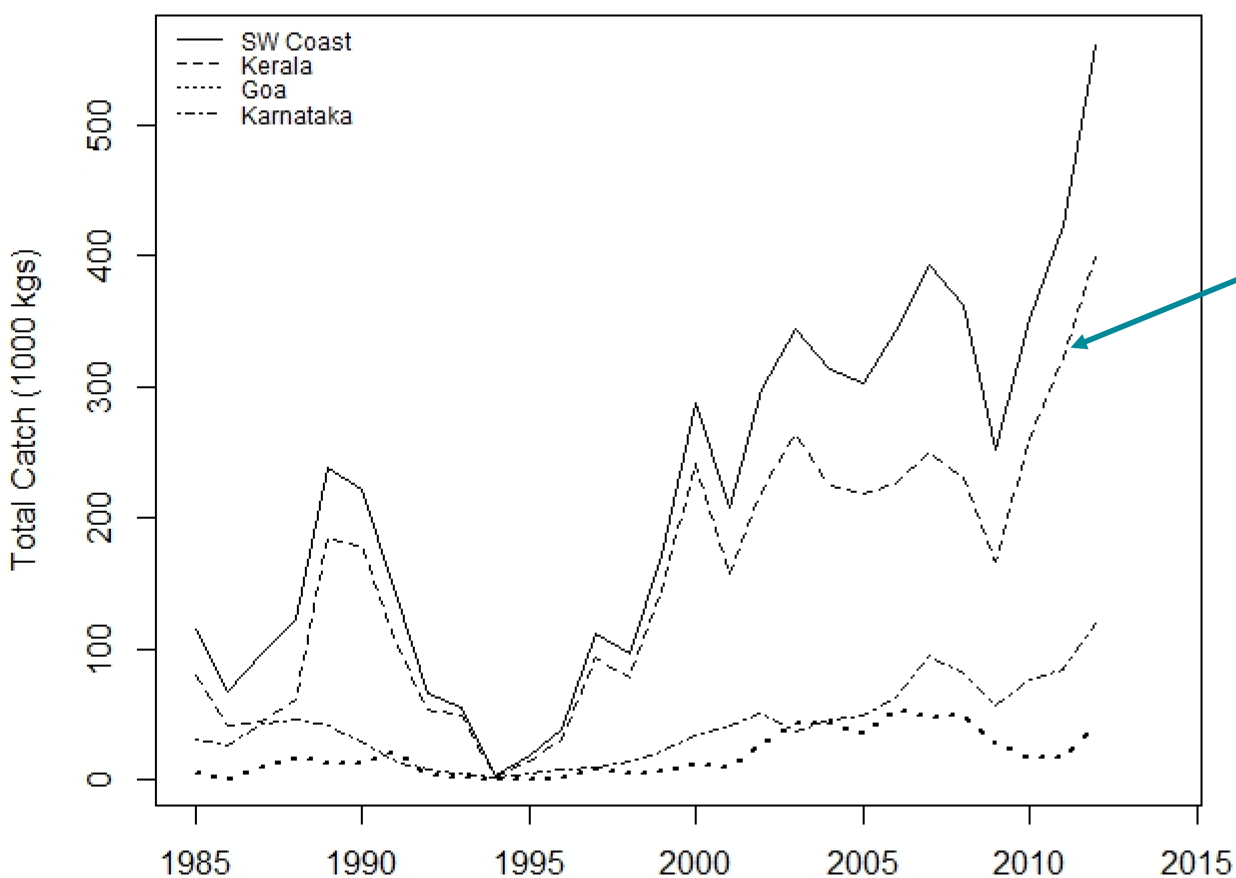
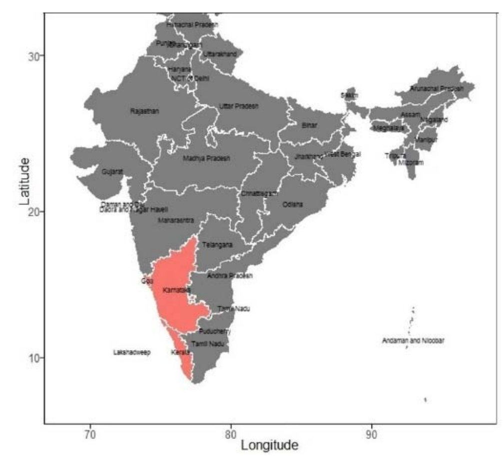


Figure 2.8 from Smitha BR, 2010, thesis

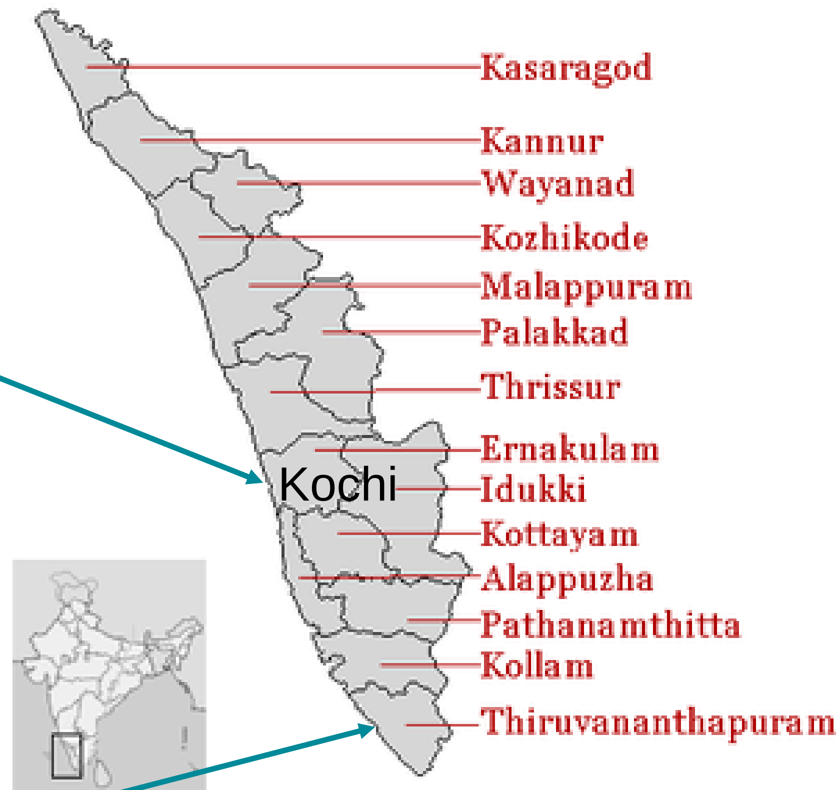
Yearly trends in the SW coast oil sardine catch since 1985



Kerala catch makes up the majority of the SW coast landings



Data collected by Central Marine Fisheries Research Institute (Kochi, Kerala)
Data on the catch by species is collected at 187 landing centers in Kerala (and over 1511 along the entire coast). This is used to produce monthly landings estimates by species.



Pictures from http://www.steyn.pro/kerala/ from a study of cell phone use by fishers


THE HINDU BusinessLine

Home News Markets Companies Money & Banking Economy Info-tech Opinion

Agri Business Logistics Macro Economy Policy

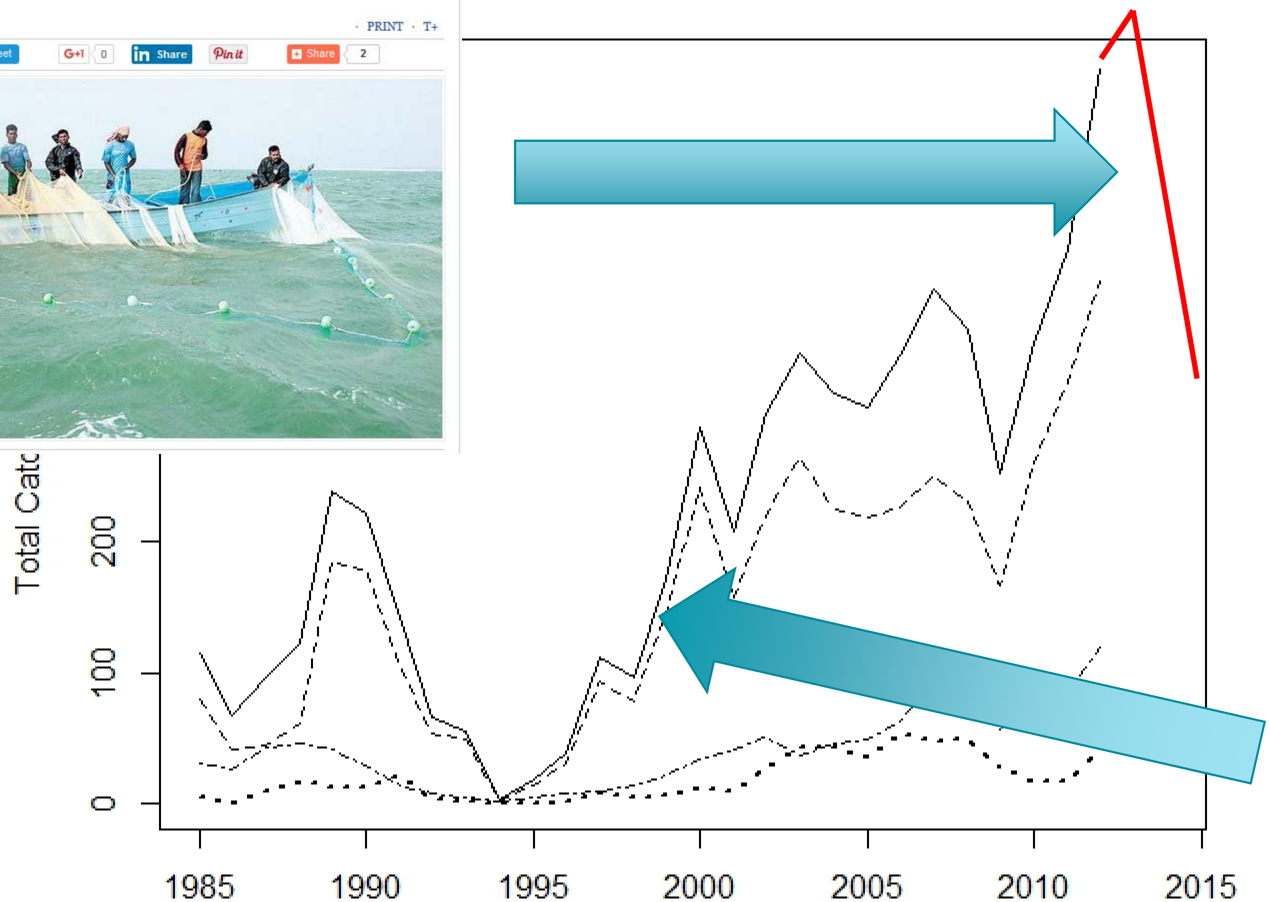
Marine fish catch drops 5.3% in 2015, 51% decline in oil sardine landings

V. SAJEEV KUMAR

PRINT · T+

Like 71 Share Tweet G+1 0 Share Pin it Share 2

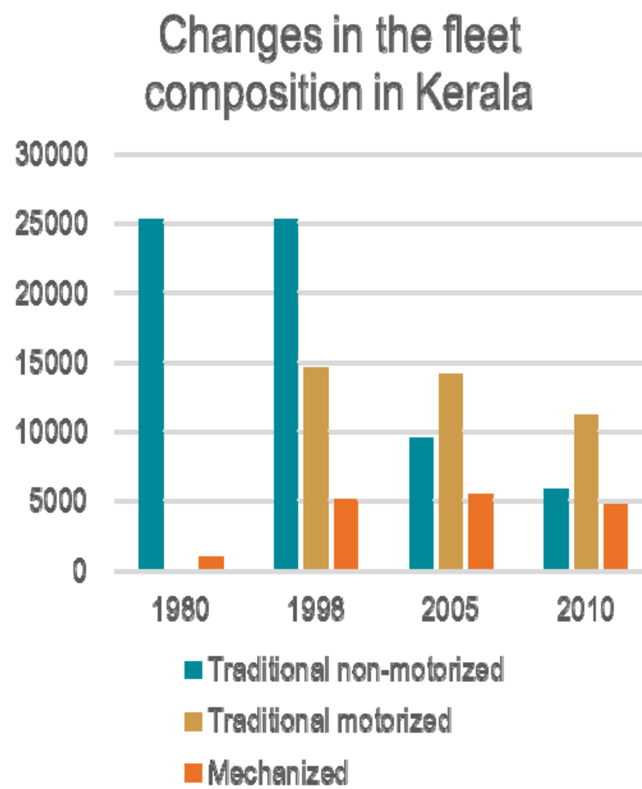
Big drop in oil sardine catch in 2015



Major change in fishery (boats)

Today oil sardines are caught primarily with ring seines using smaller motorized (outboard) boats or larger (inboard) ships

(The Hindu)



1985-2014 Changes in the ring seine boats and gear

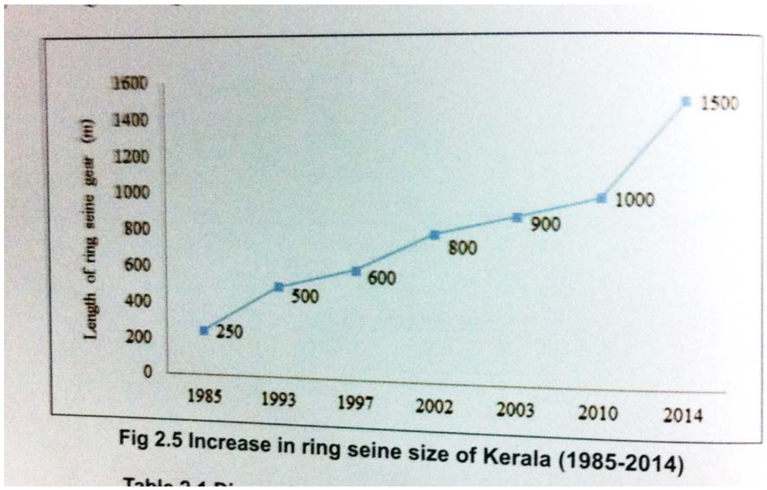
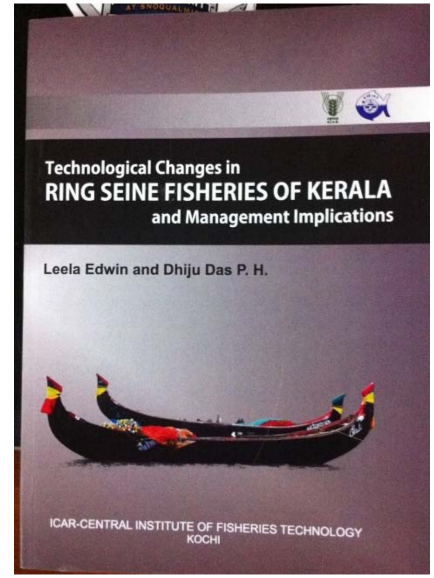


Fig 2.5 Increase in ring seine size of Kerala (1985-2014)

Nets are getting longer



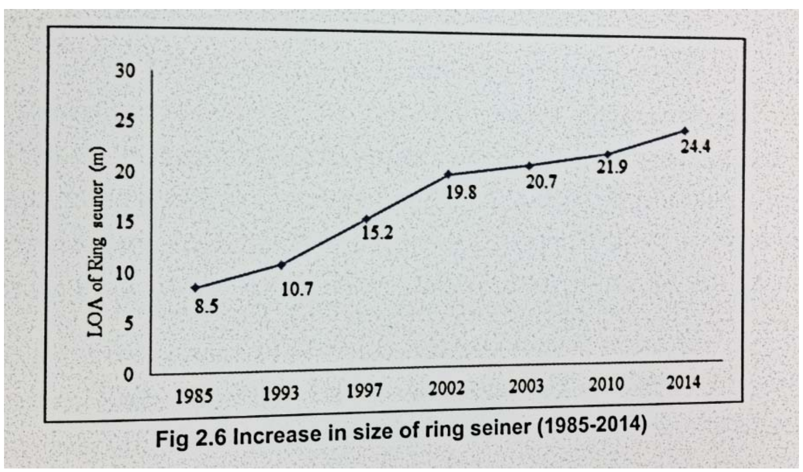


Fig 2.6 Increase in size of ring seiner (1985-2014)

Boats are getting longer

Engines are getting stronger

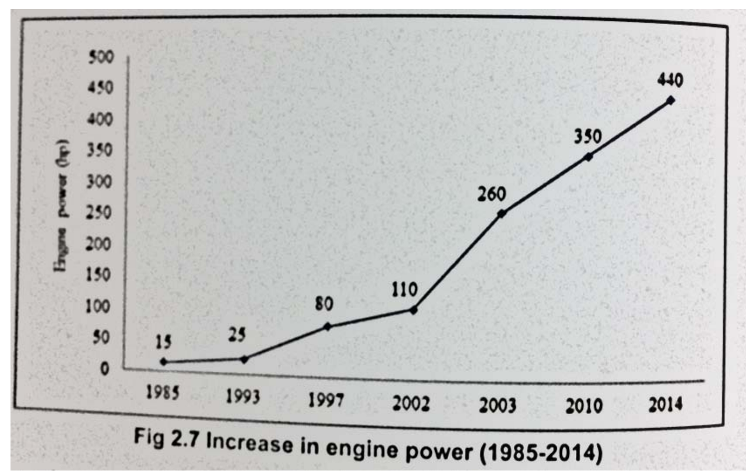
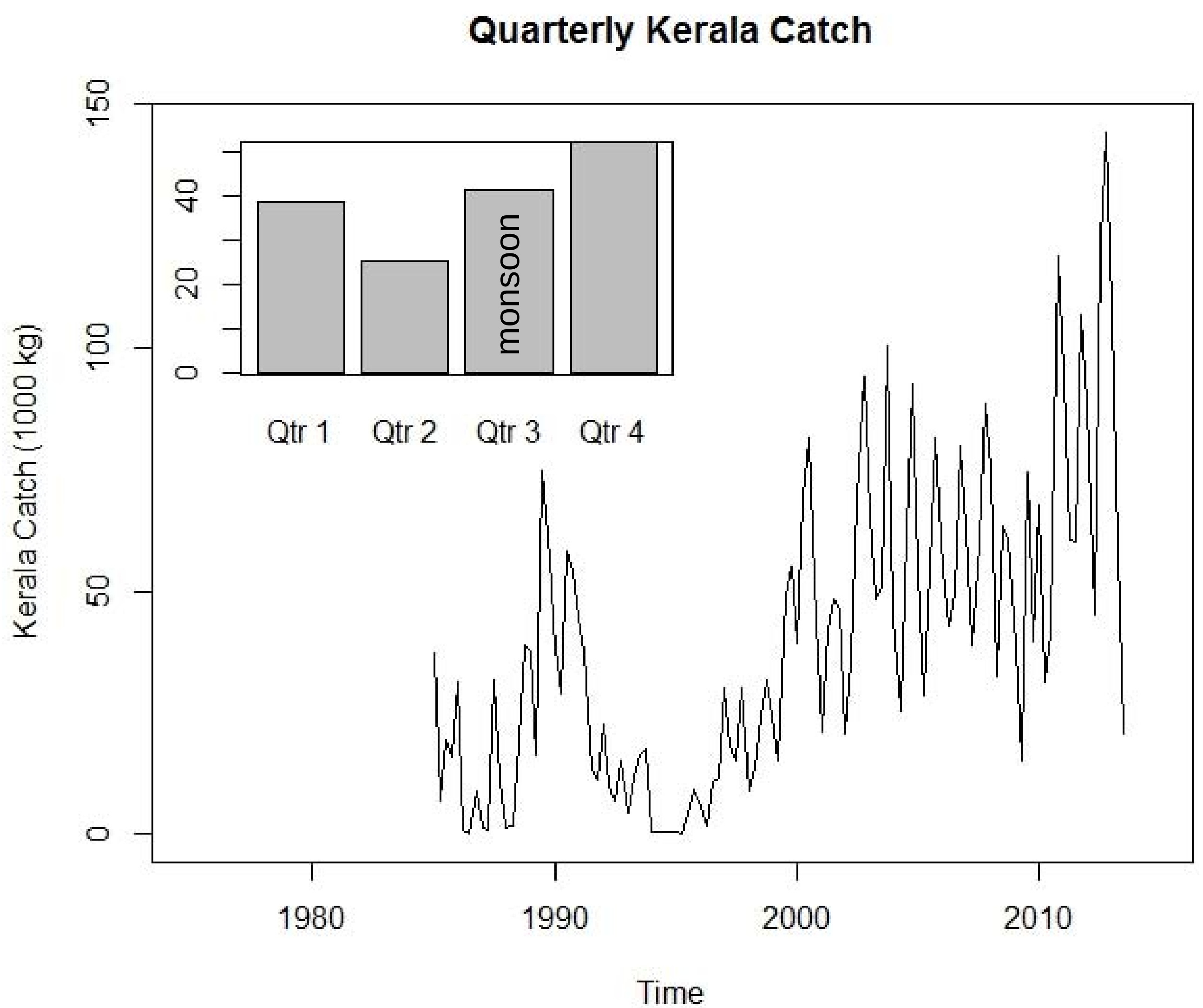
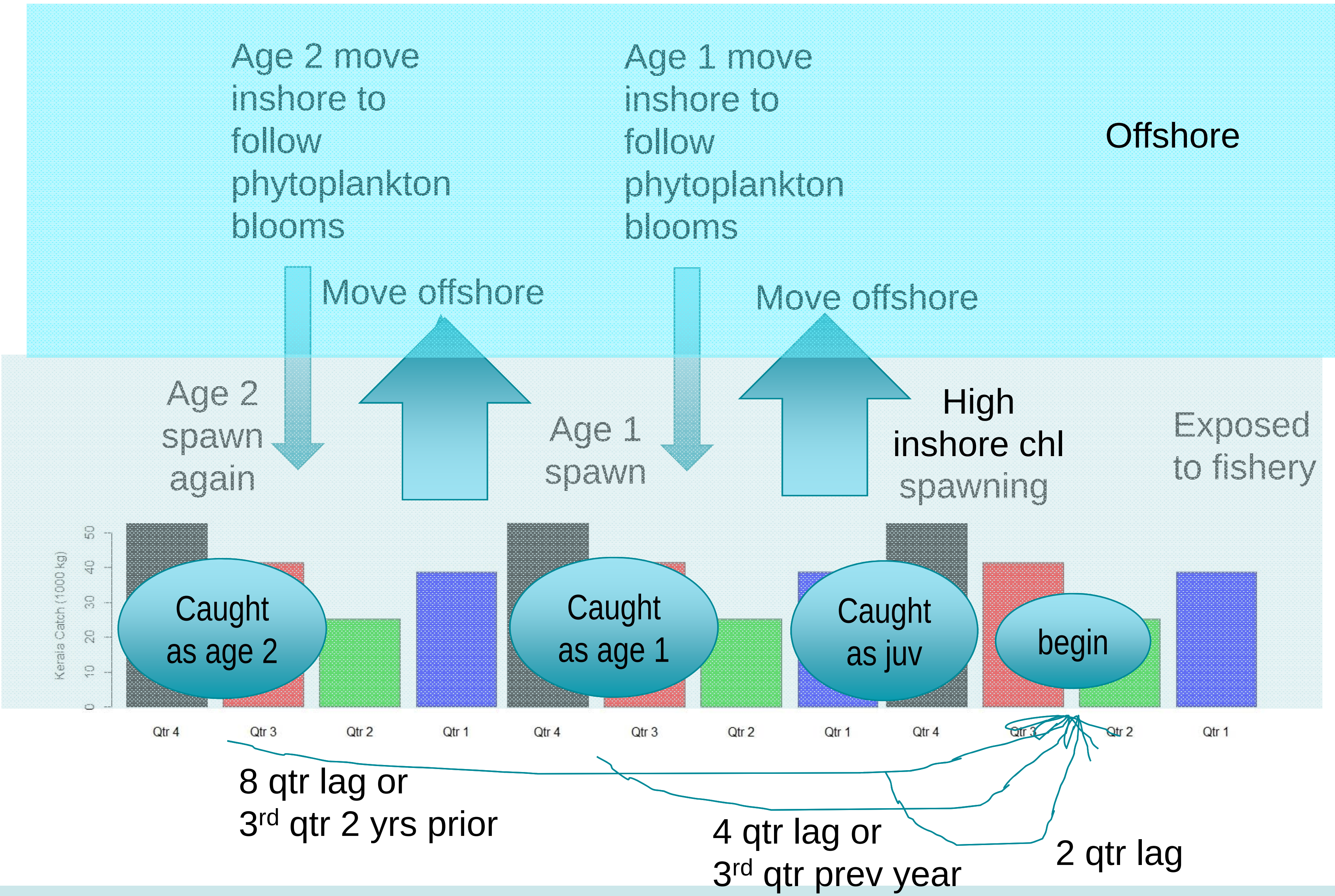


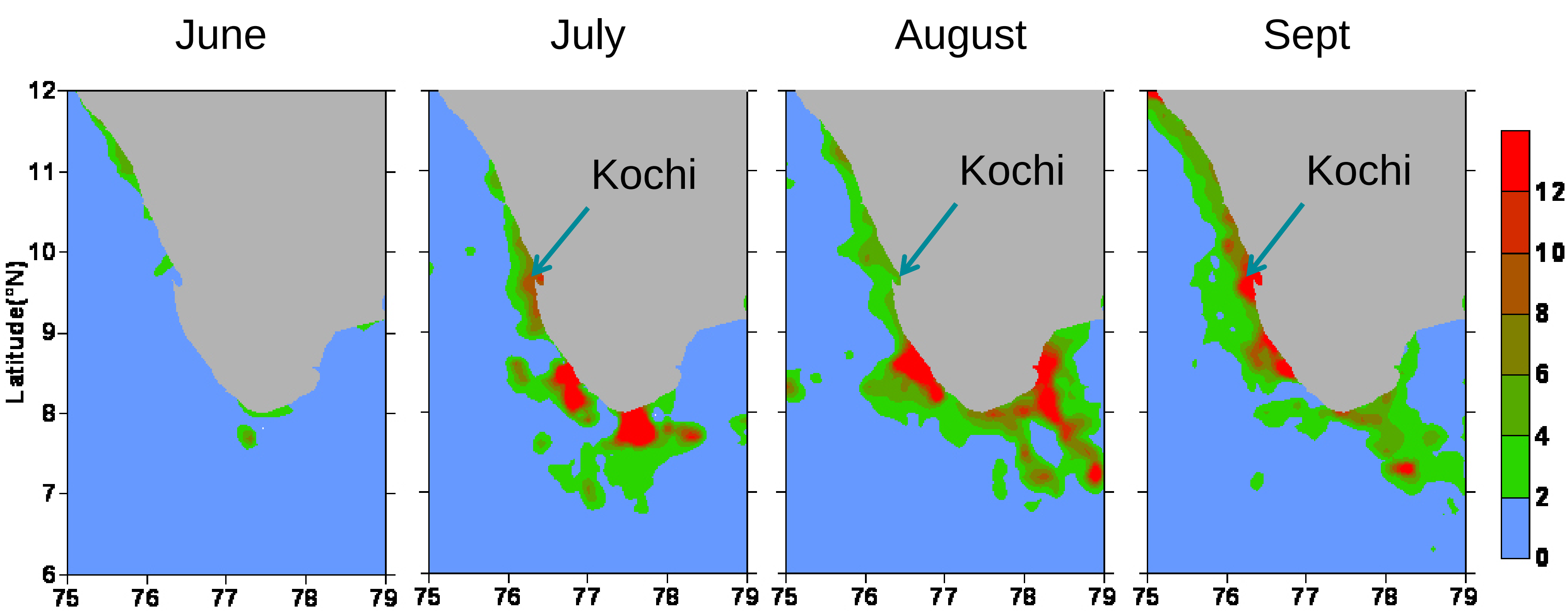
Fig 2.7 Increase in engine power (1985-2014)


Quarterly Kerala Catch
Kerala Catch (1000 kg)
0
50
100
150
0
20
40
monsoon
Qtr 1
Qtr 2
Qtr 3
Qtr 4
1980
1990
2000
2010
Time



Age 2 move inshore to follow phytoplankton blooms
Age 1 move inshore to follow phytoplankton blooms
Offshore
Move offshore
Move offshore
Age 2 spawn again
Age 1 spawn
High inshore chl
spawning
Exposed to fishery
Keraia Catch (1000 kg)
0
10
20
30
40
50
Caught as age 2
Caught as age 1
Caught as juv
begin
Qtr 4
Qtr 3
Qtr 2
Qtr 1
Qtr 4
Qtr 3
Qtr 2
Qtr 1
Qtr 4
Qtr 3
Qtr 2
Qtr 1
8 qtr lag or 3rd qtr 2 yrs prior
4 qtr lag or 3rd qtr prev year
2 qtr lag

Upwelling indices



Four-year (2003–2006) average monthly monthly surface Chlorophyll from MODIS AQUA for June–September.(monsoon months)

Figure 2.8 from Smitha BR, 2010, thesis

Upwelling indices

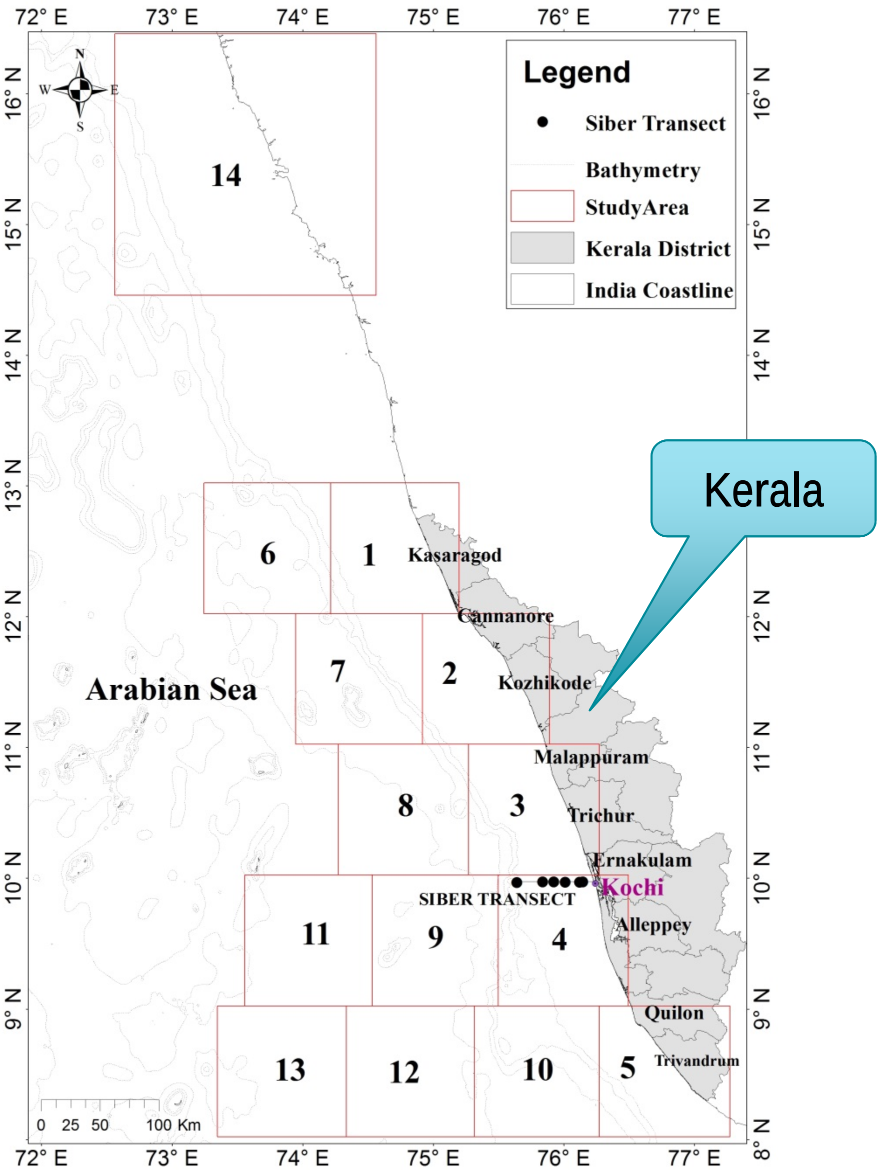
- Surface SST
  - Upwelling brings cooler water to the surface
  - The offshore versus onshore differential is used as an upwelling index
- SSH anomaly
  - Upwelling is associated with a drop in SSH
- Surface Chlorophyll
- Wind speed derived upwelling index: BR Smitha, 2010

Data

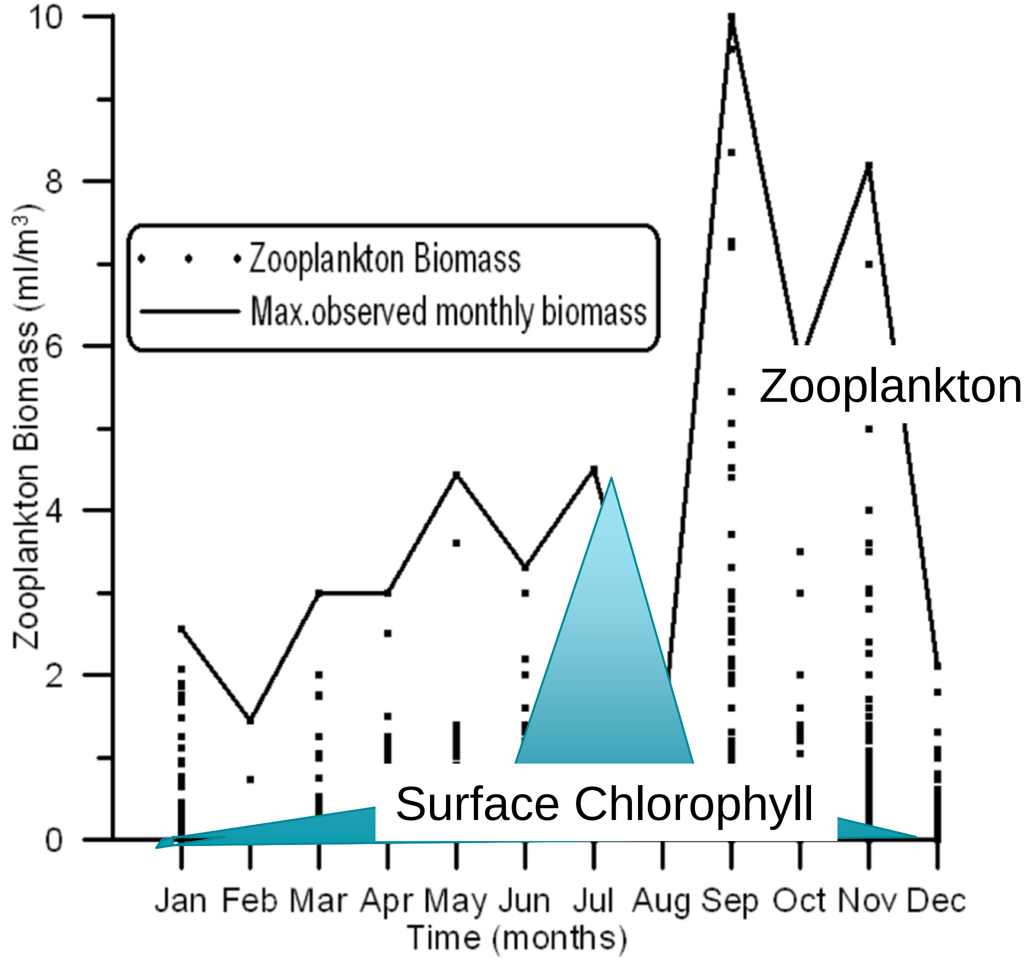
Landing data
1985-2015 quarterly *landings*
Kerala, Kernataka, Goa

Satellite data
SST
SSH
Chl-a
Precipitation

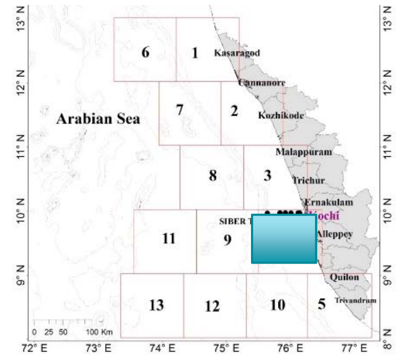
Other
10-yr Upwelling index (wind-derived)
Local measurements (DO, salinity, zooplankton)
Monsoon onset dates

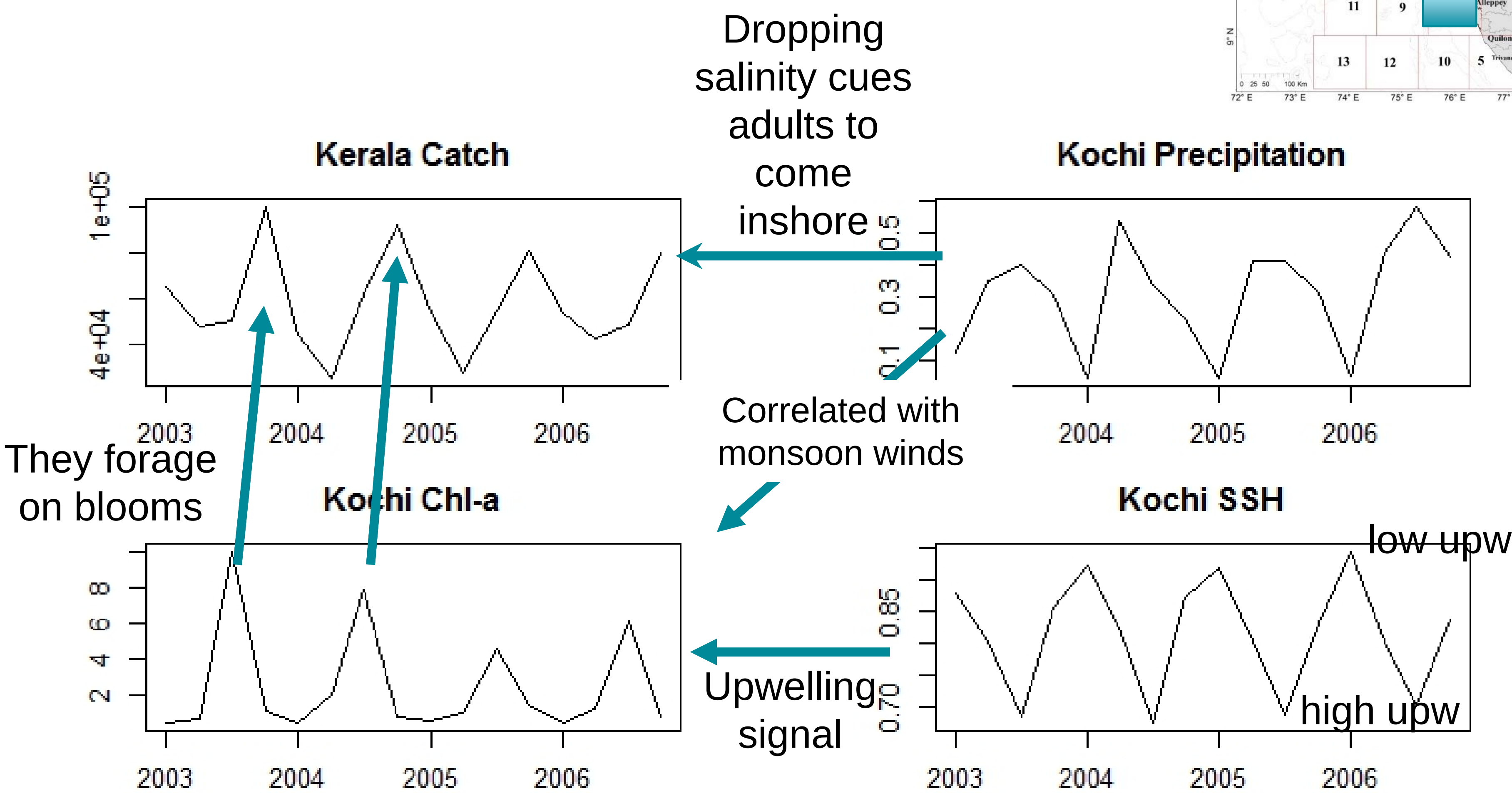


We expect lagged relationships between our upwelling covariates, Chl, and catch



Seasonality



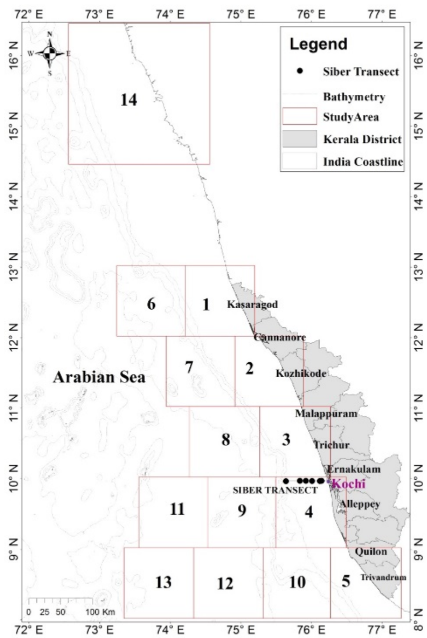


High food = higher survival of juveniles

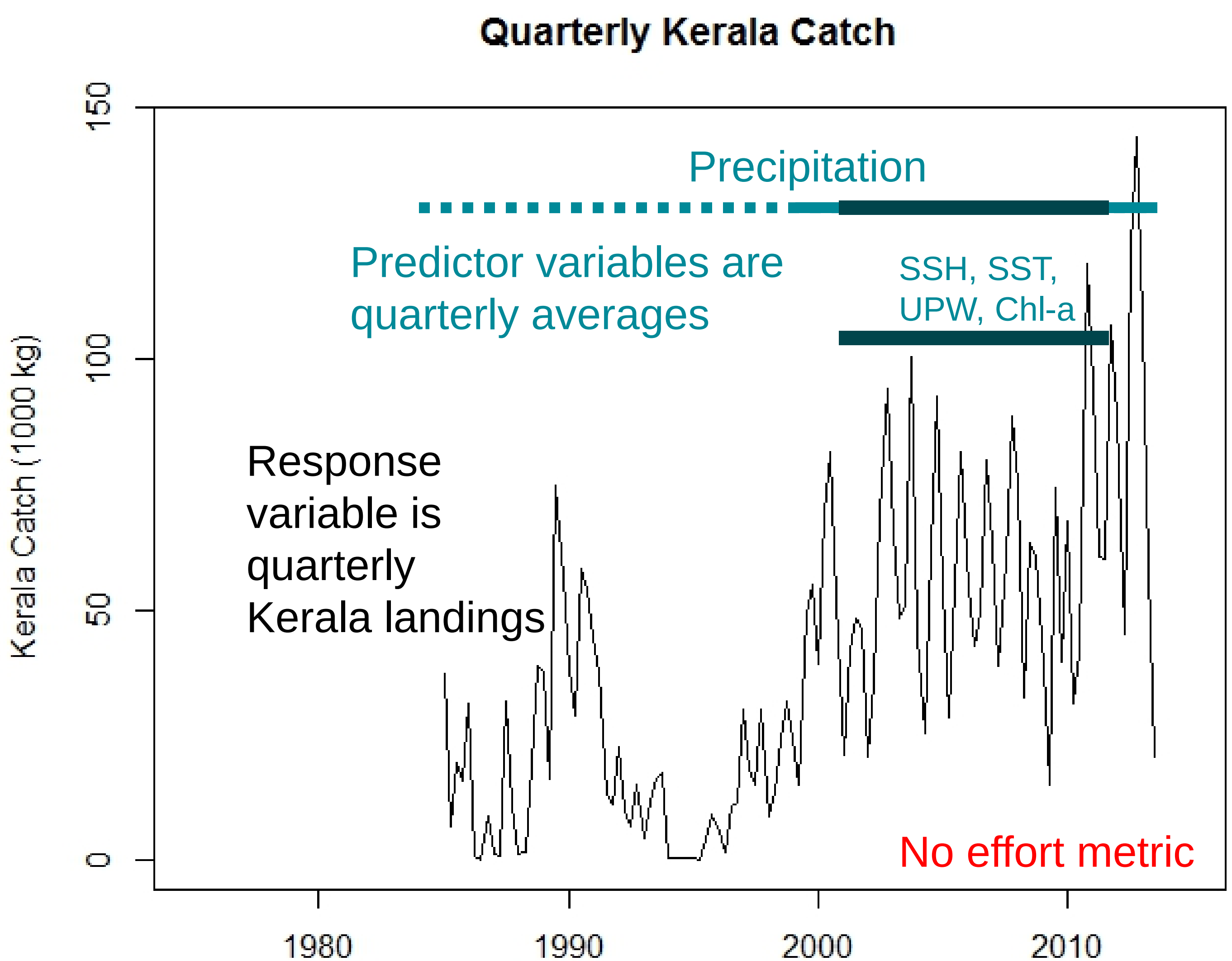
Statistical Analysis

- What covariates (if any) correlate with landings?
  - 13 boxes, 5 covariates
  - ca 40 landing data points: 2002-2012
- Start with looking at the correlations

I used cross-correlation and looking at the R2 for various multivariate regressions, and then step-AIC regressions. In the end, looking at simple plots seemed most informative.

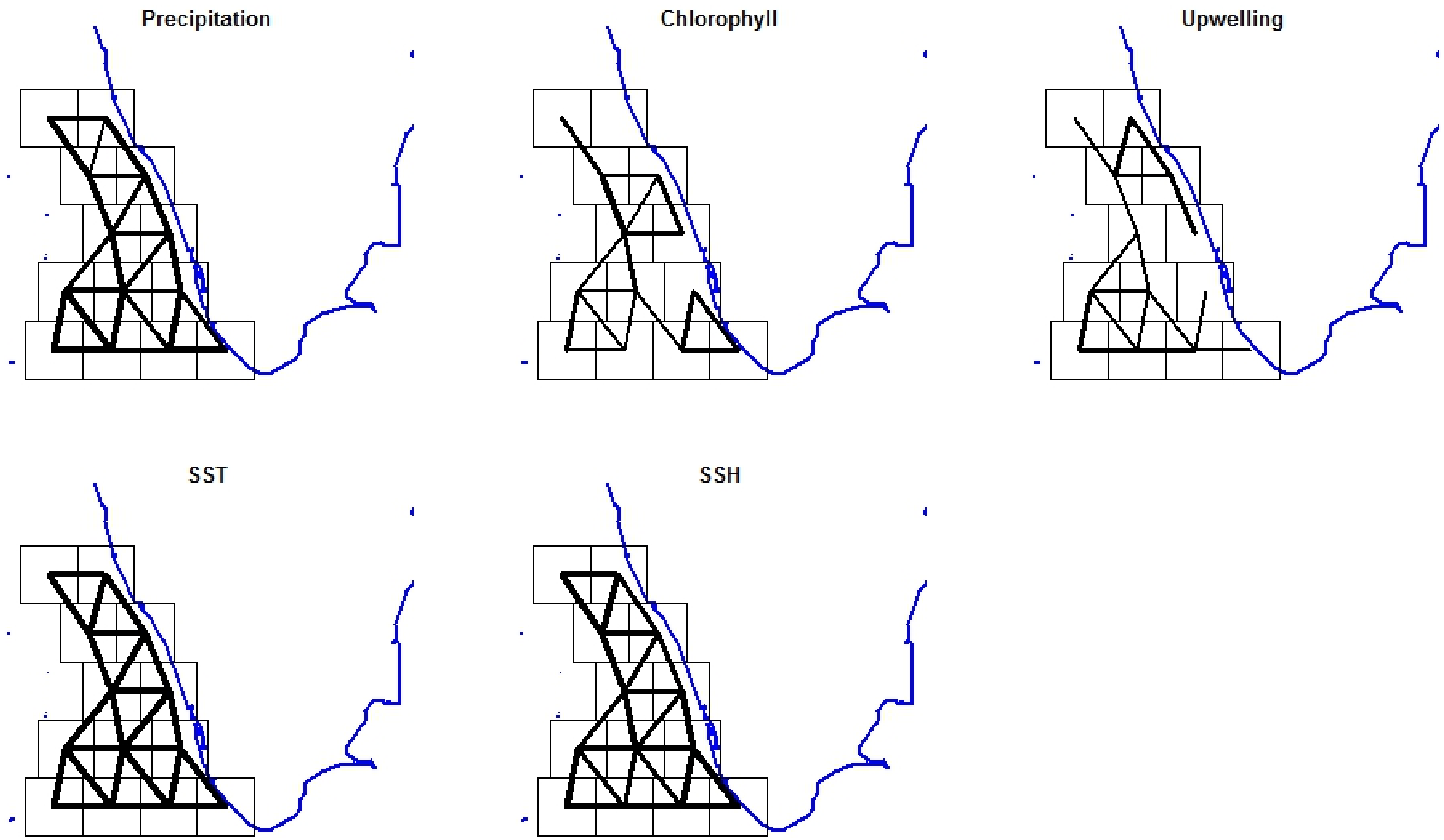


Data used

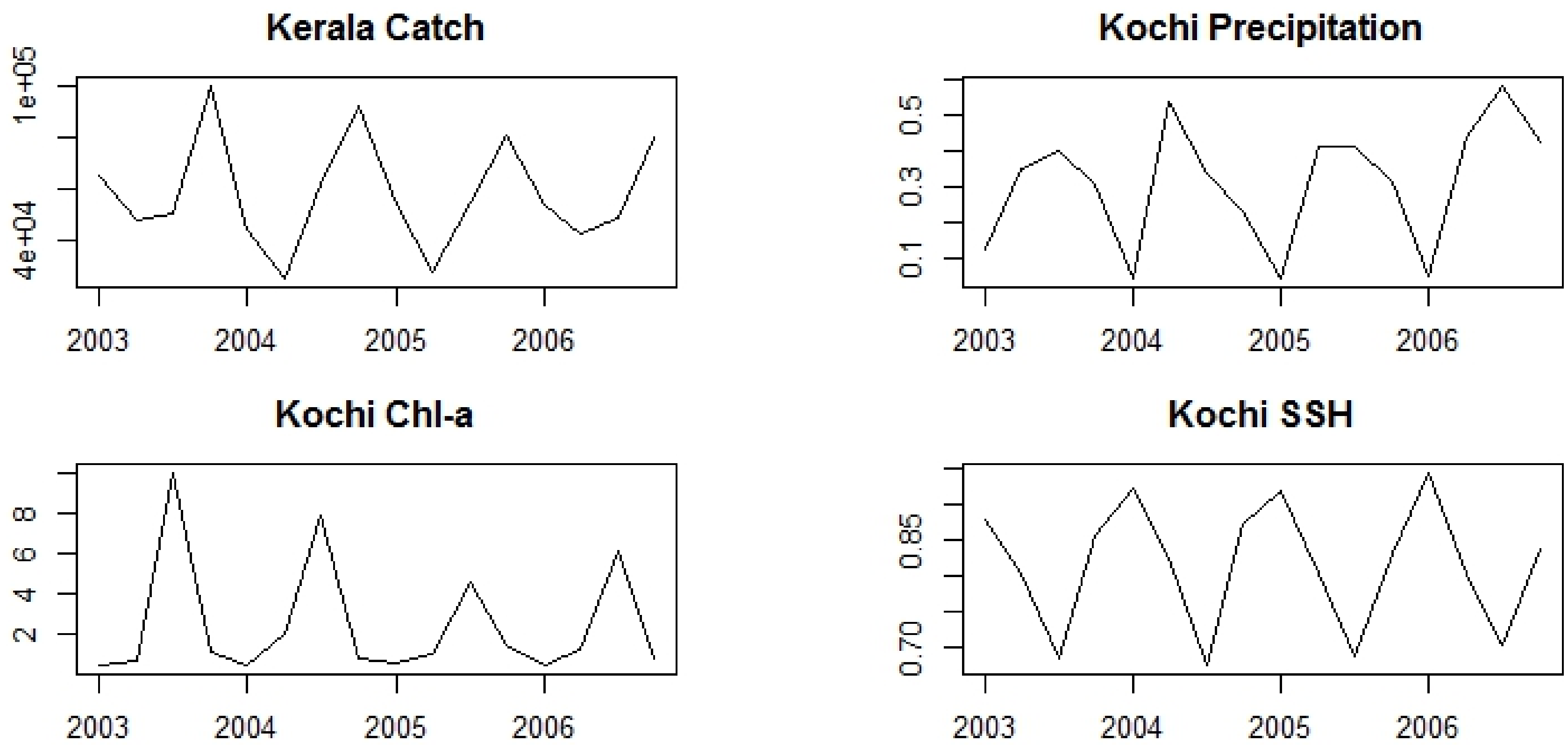


Quarterly Kerala Catch
Kerala Catch (1000 kg)
150
100
50
0
1980
1990
2000
2010
Precipitation
Predictor variables are
quarterly averages
SSH, SST,
UPW, Chl-a
Response
variable is
quarterly
Kerala landings
No effort metric

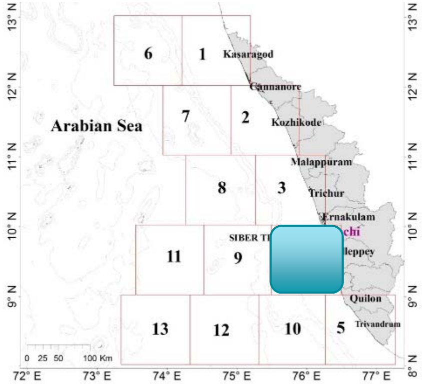
The covariates are spatially correlated
Which location(s) is(are) important?

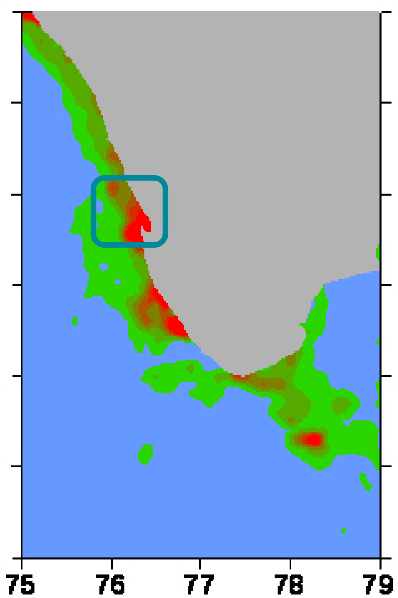


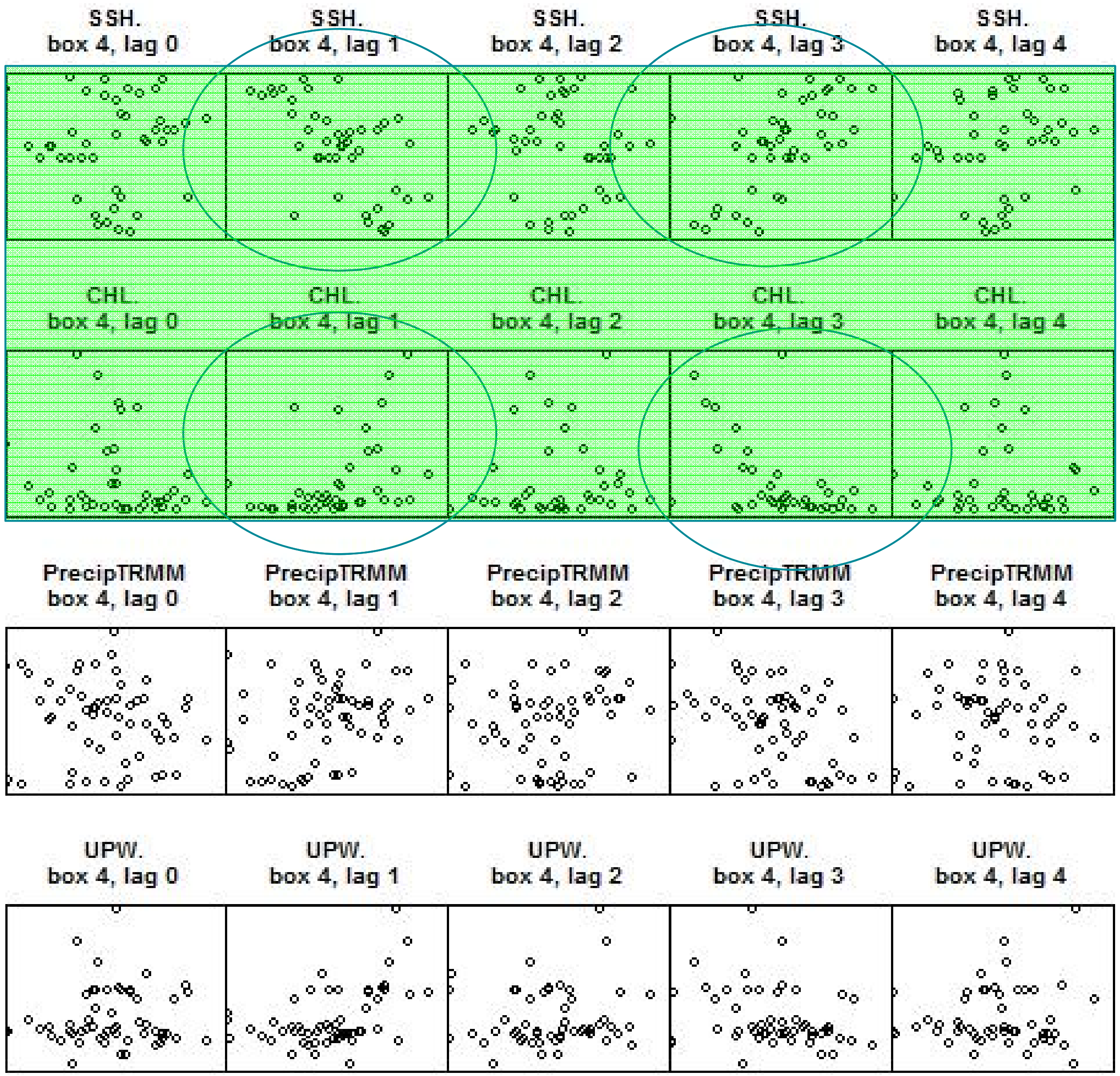
The covariates are temporally correlated. *Which covariate is a good predictor of the seasonal cycle in catch?*

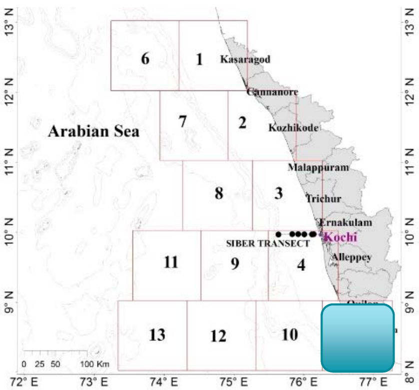


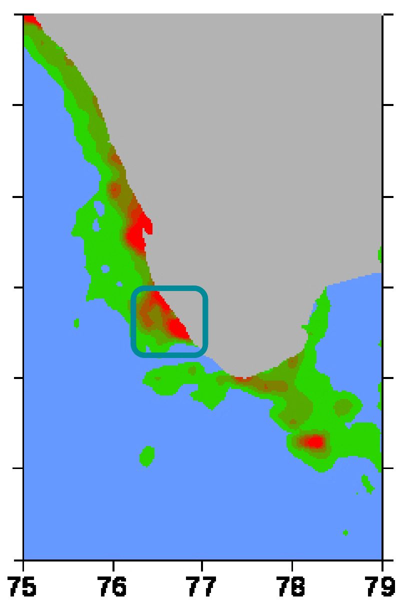
Box 4 off Kochi

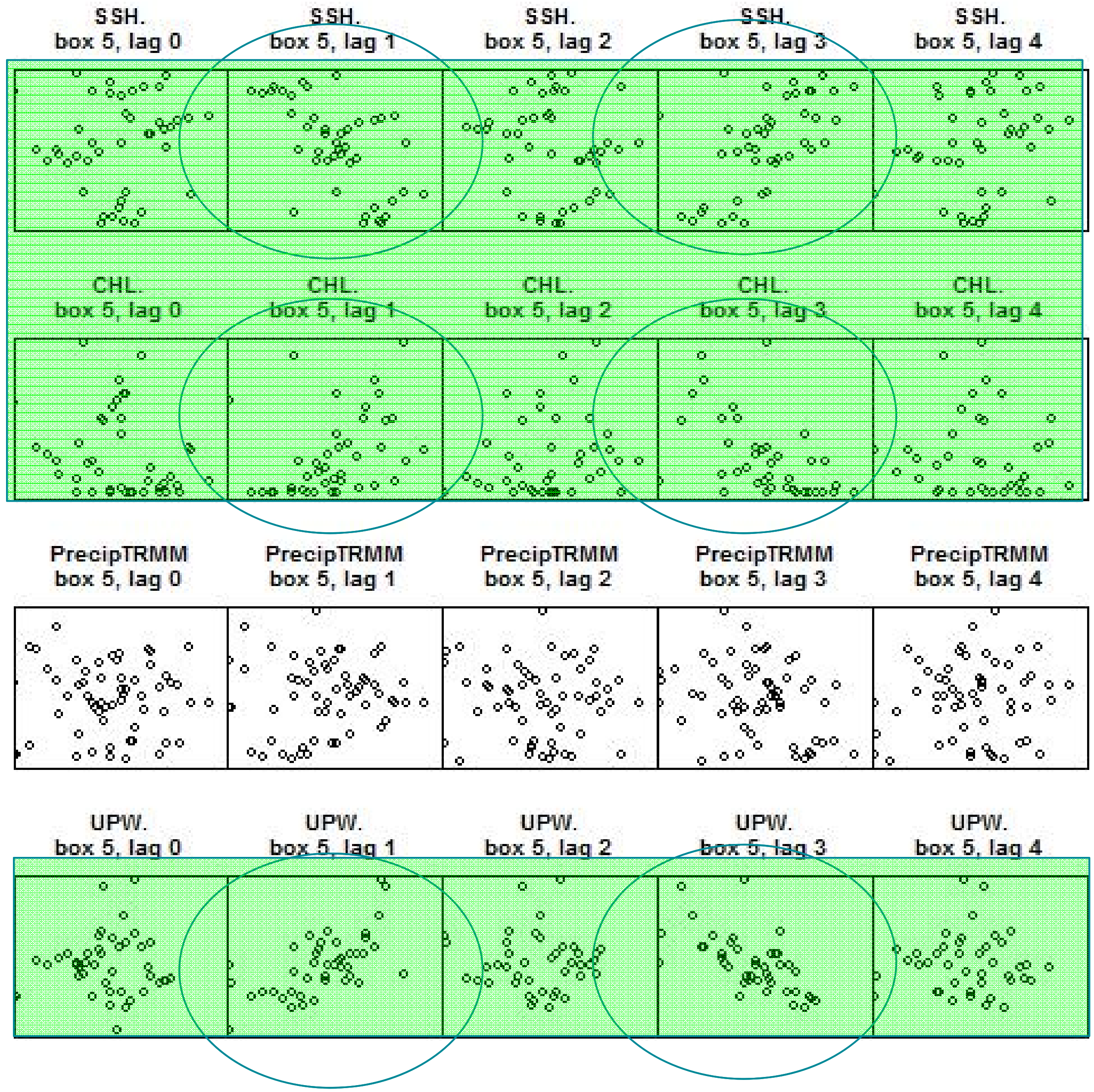

Arabian Sea
Kasaragod
Cannanore
Kozhikode
Malappuram
Trichur
Ernakulam
Quilon
Trivandrum
1
2
3
5
6
7
8
9
10
11
12
13
72° E
73° E
74° E
75° E
76° E
77° E
8° N
9° N
10° N
11° N
12° N
13° N

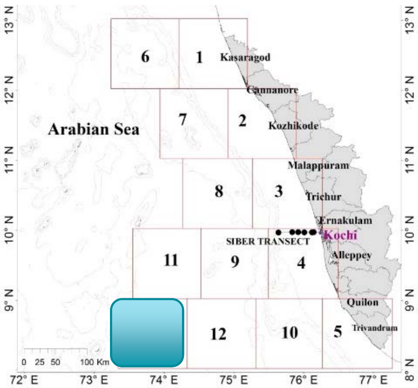

75
76
77
78
79


SSH.
box 4, lag 0
SSH.
box 4, lag 1
SSH.
box 4, lag 2
SSH.
box 4, lag 3
SSH.
box 4, lag 4
CHL.
box 4, lag 0
CHL.
box 4, lag 1
CHL.
box 4, lag 2
CHL.
box 4, lag 3
CHL.
box 4, lag 4
PrecipTRMM
box 4, lag 0
PrecipTRMM
box 4, lag 1
PrecipTRMM
box 4, lag 2
PrecipTRMM
box 4, lag 3
PrecipTRMM
box 4, lag 4
UPW.
box 4, lag 0
UPW.
box 4, lag 1
UPW.
box 4, lag 2
UPW.
box 4, lag 3
UPW.
box 4, lag 4

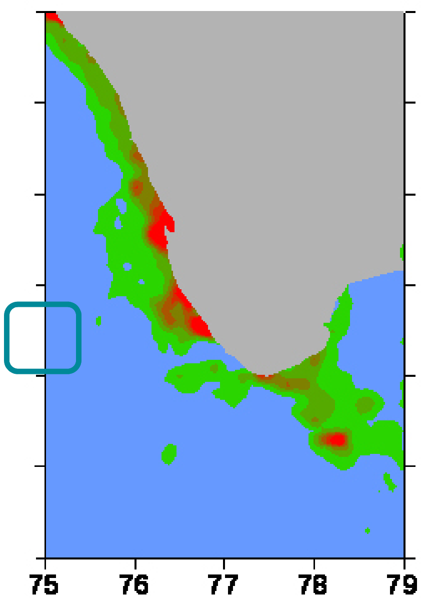

13° N
12° N
11° N
10° N
9° N
8° N
6
1
Kasaragod
Cannanore
7
2
Kozhikode
Arabian Sea
Malappuram
8
3
Trichur
Ernakulam
Kochi
SIBER TRANSECT
11
9
4
Alleppey
13
12
10
0 25 50 100 Km
72° E
73° E
74° E
75° E
76° E
77° E


75
76
77
78
79

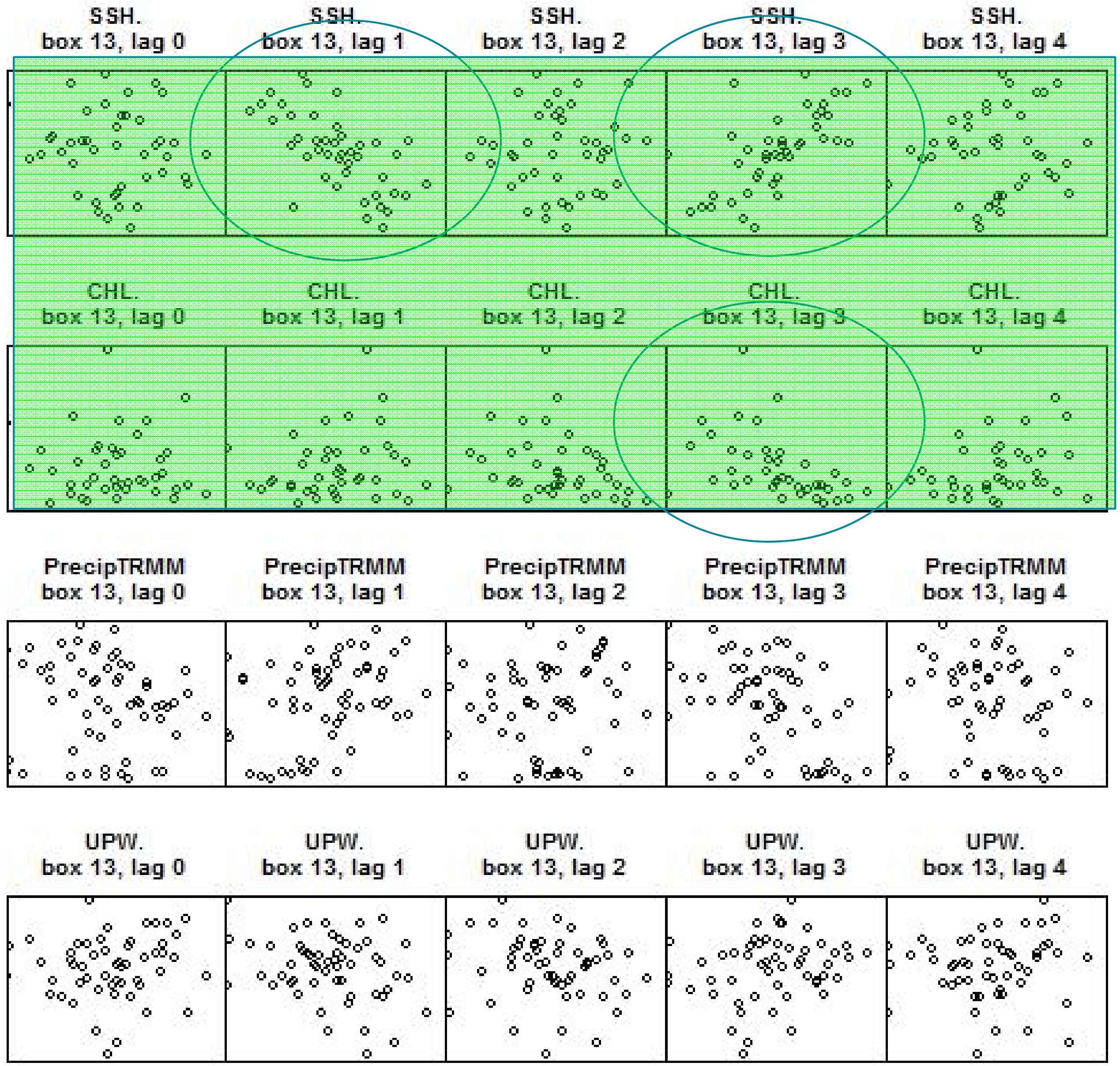

SSH.
box 5, lag 0
SSH.
box 5, lag 1
SSH.
box 5, lag 2
SSH.
box 5, lag 3
SSH.
box 5, lag 4
CHL.
box 5, lag 0
CHL.
box 5, lag 1
CHL.
box 5, lag 2
CHL.
box 5, lag 3
CHL.
box 5, lag 4
PrecipTRMM
box 5, lag 0
PrecipTRMM
box 5, lag 1
PrecipTRMM
box 5, lag 2
PrecipTRMM
box 5, lag 3
PrecipTRMM
box 5, lag 4
UPW.
box 5, lag 0
UPW.
box 5, lag 1
UPW.
box 5, lag 2
UPW.
box 5, lag 3
UPW.
box 5, lag 4



Arabian Sea
6
1
Kasaragod
Cannanore
7
2
Kozhikode
Malappuram
8
3
Trichur
Ernakulam
Kochi
SIBER TRANSECT
11
9
4
Alleppey
Quilon
12
10
5
Trivandrum
72° E
73° E
74° E
75° E
76° E
77° E
13° N
12° N
11° N
10° N
9° N
8° N



75
76
77
78
79



SSH.
box 13, lag 0
SSH.
box 13, lag 1
SSH.
box 13, lag 2
SSH.
box 13, lag 3
SSH.
box 13, lag 4
CHL.
box 13, lag 0
CHL.
box 13, lag 1
CHL.
box 13, lag 2
CHL.
box 13, lag 3
CHL.
box 13, lag 4
PrecipTRMM
box 13, lag 0
PrecipTRMM
box 13, lag 1
PrecipTRMM
box 13, lag 2
PrecipTRMM
box 13, lag 3
PrecipTRMM
box 13, lag 4
UPW.
box 13, lag 0
UPW.
box 13, lag 1
UPW.
box 13, lag 2
UPW.
box 13, lag 3
UPW.
box 13, lag 4

What covariates are promising for modeling seasonality?

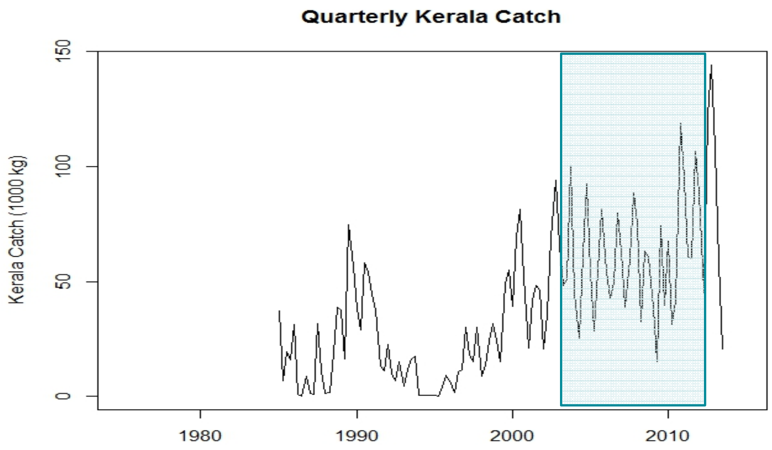
- SSH at lag 1 or lag 3 (prev qtr or 3 qtr prior). No region jumped out as better than another probably because SSH is highly spatially correlated.
- The wind-based upwelling index in the south is also a contender.
- Chl-a isn't as good of a predictor though the correlation is in the right direction. My Chl-a covariate is probably too crude (average quarterly). Others have used the peak (not the average) with success.
- Precipitation was not predictive, too variable. Maybe using average is poor. Biologists think that the date of monsoon onset is important.

But quarter works just fine for modeling seasonality….

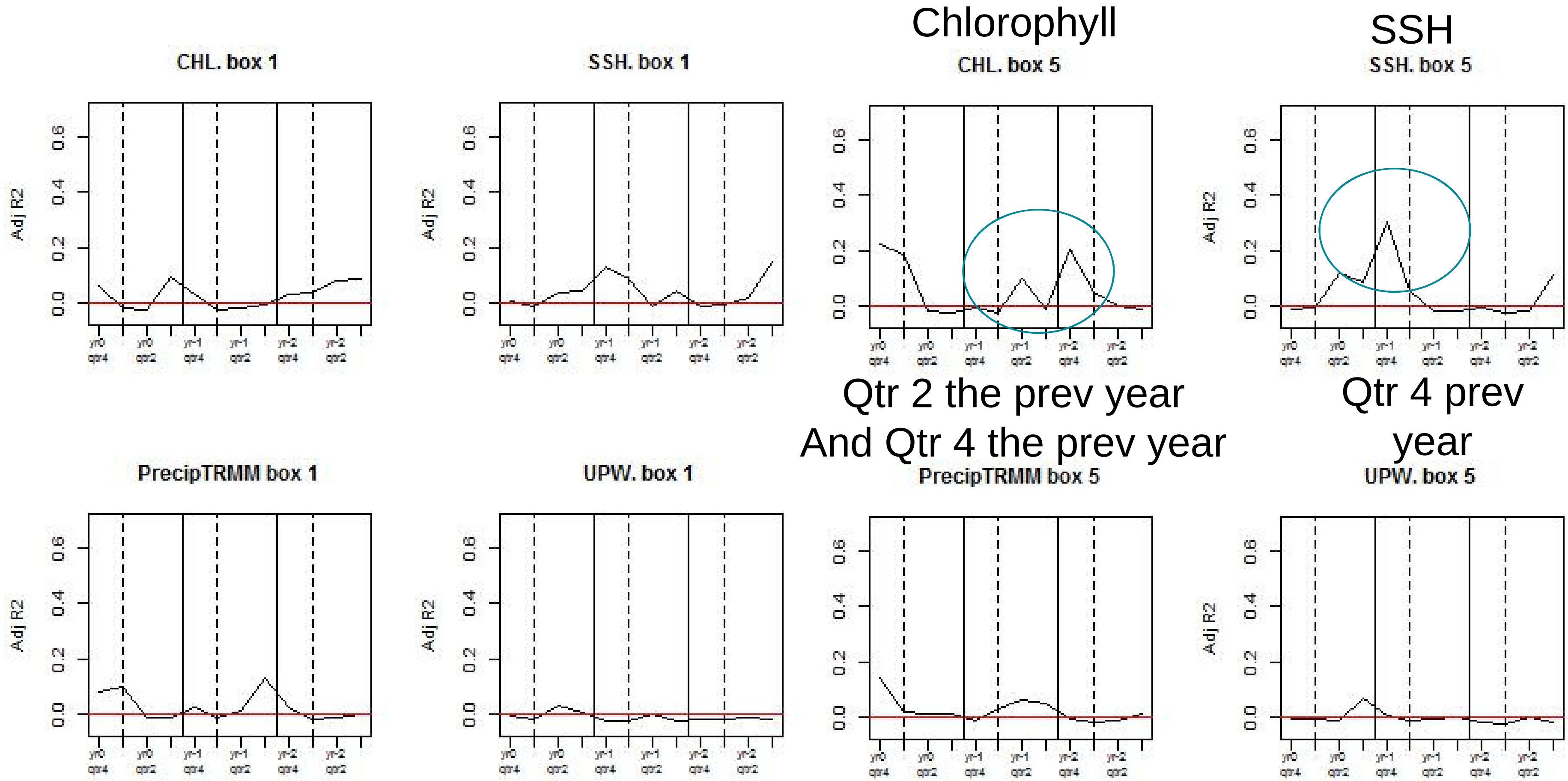
- Just using Quarter and Year is better for years that I have for testing.

Log catch ~ Year + Qtr	adj R2 = 0.52
Log catch ~ Year + Chl(1 qtr prev)	adj R2 = 0.26
Log catch ~ Year + Chl(3 qtrs. prev)	adj R2 = 0.43
Log catch ~ Year + SSH(1 qtr prev)	adj R2 = 0.52
Log catch ~ Year + SSH(3 qtrs. prev)	adj R2 = 0.49

2003-2012
40 data points



What about predicting catch anomalies? Effect of covariate anomaly in a particular qtr.



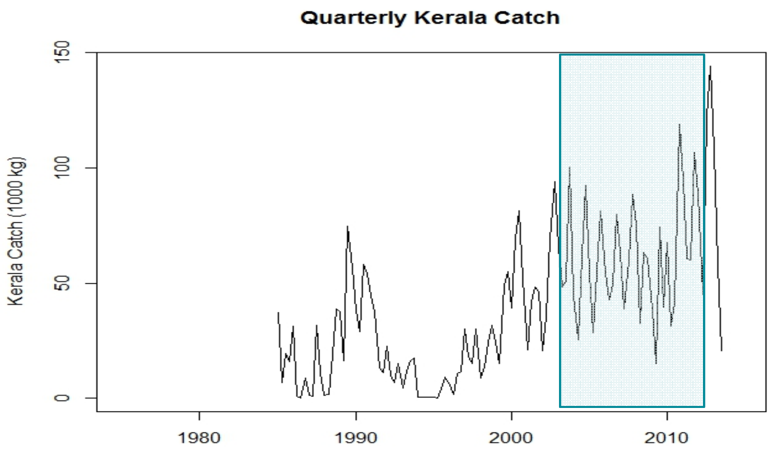
Using covariate anomalies seems promising

Log catch ~ Year + Qtr adj R2 = 0.55

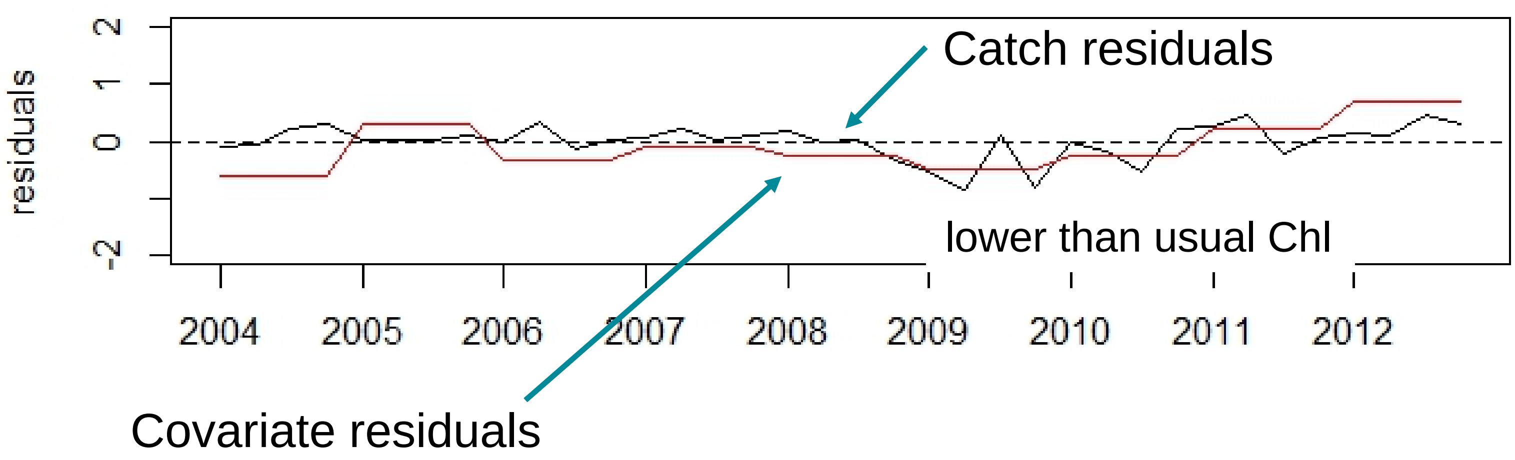
Log catch ~ Year + Qtr + Chl anomaly(qtr 2 prior year) adj R2 = 0.65

Log catch ~ Year + Qtr + SSH anomaly(qtr 4 prior year) adj R2 = 0.66

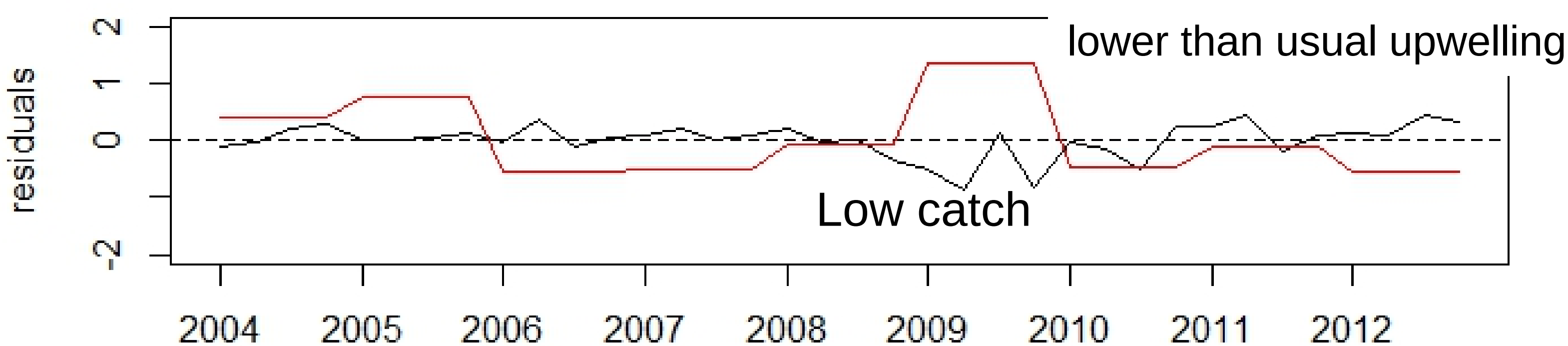
2004-2012
36 data points



CHL in box 5 during qtr 2 (pre-monsoon)
only boxes 4 and 5 work



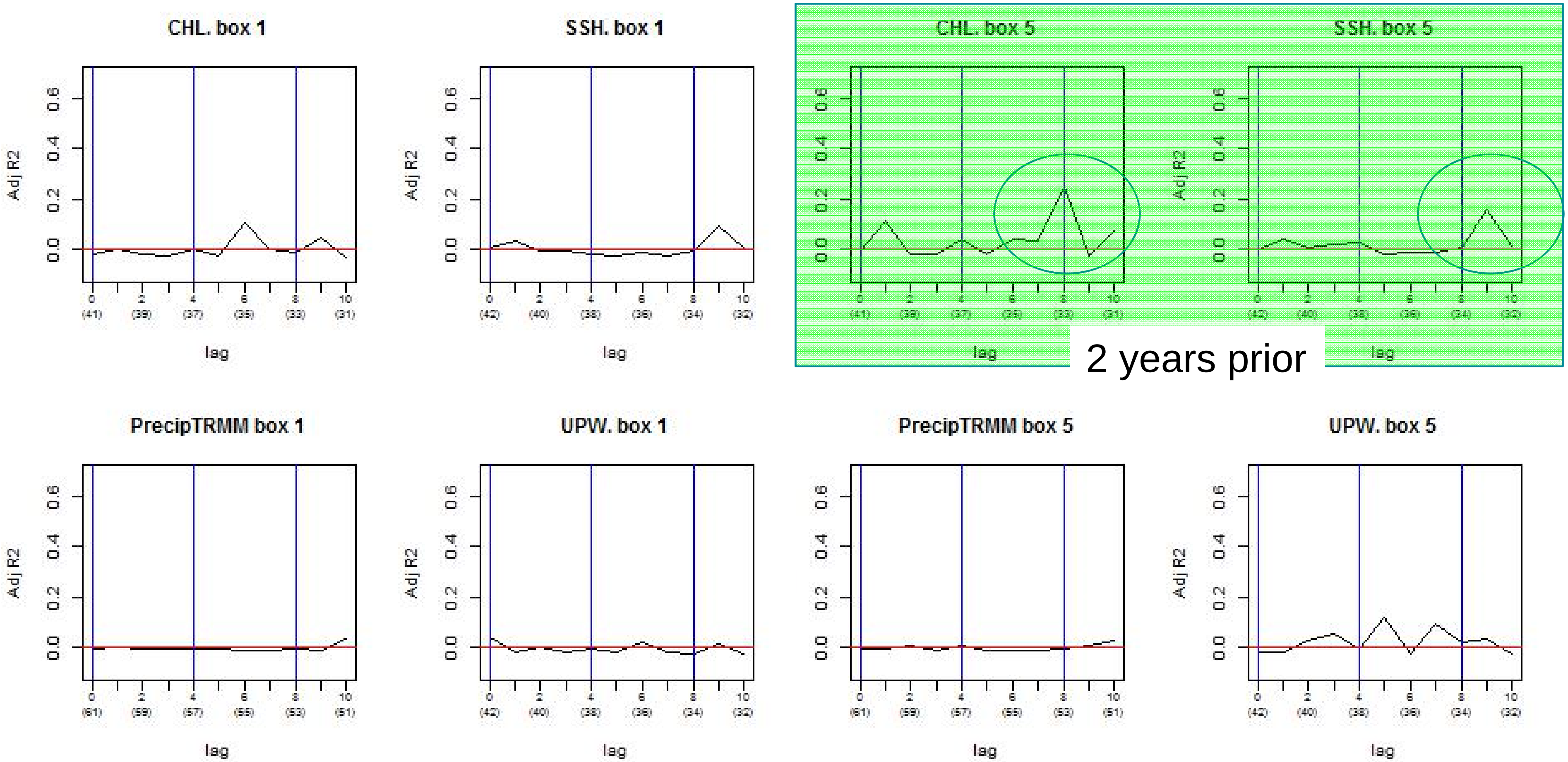
SSH in box 5 during qtr 4 (post-monsoon)
any inshore box works similarly



Box 4 not box 5

Does predicting catch anomalies using lags work?

Effect of covariate at a particular lag (# qtrs. In the past)



Today's talk --- overview of the project, challenges, and initial results

- Background
  - Physical processes (upwelling) affecting oil sardines in the SE Arabian Sea
  - Biology of oil sardines and how the interacts with the above
- Satellites covariates
- Exploratory correlation analysis
  - Which covariates have explanatory value? biology + physical processes + information in the covariate
- Testing some forecasting models
  - Multivariate regression models
  - Exponential smoothing models

Multiple regression model

$$y_t = \text{level} + \text{qtr}_t + b1 * \text{cov1}_t + b2 * \text{cov2}_t + e_t$$

y_t is the log catch at t

Level (intercept) is constant

qtr_t is a factor for the season; qtr factor is constant

b1 and b2 are constant (not time-varying)

e_t is i.i.d. and normal

Exponential smoothing model

Time-series model that allows a time-varying level and time-varying seasonal amplitude

data | Prediction = level + season | error

$$y_t = l_{t-1} + s_{t+m} + e_t$$

$$l_t = l_{t-1} + \alpha e_t$$

Level is a random walk

$$s_t = s_{t-m} + \gamma e_t$$

M is the period of the season
Season is a random walk

Only 1 error at each t

Forecast library in R easily fits these

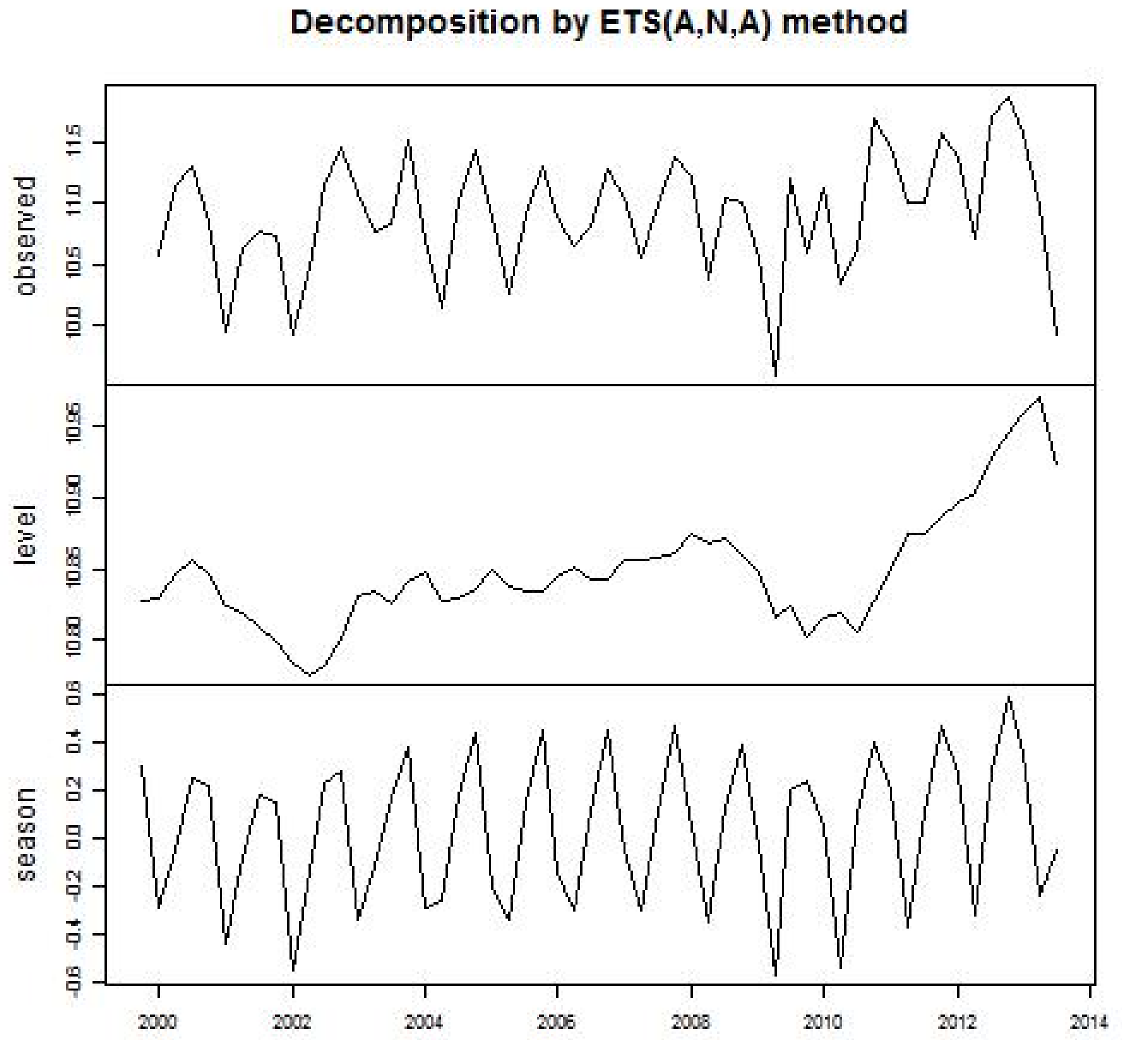
Exponential smoothing model output

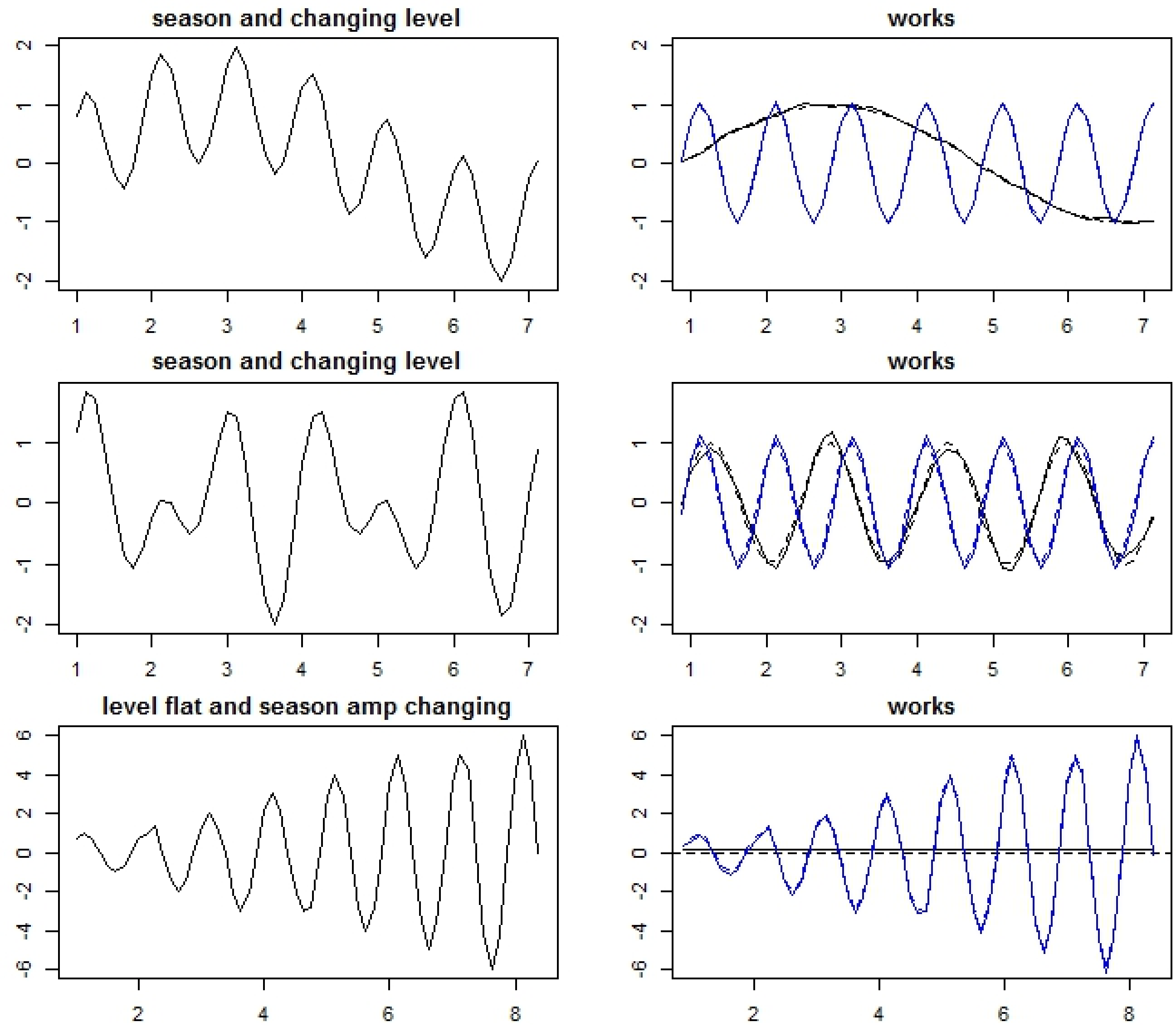
mod=ets(data, model="ANA"); fit(mod)

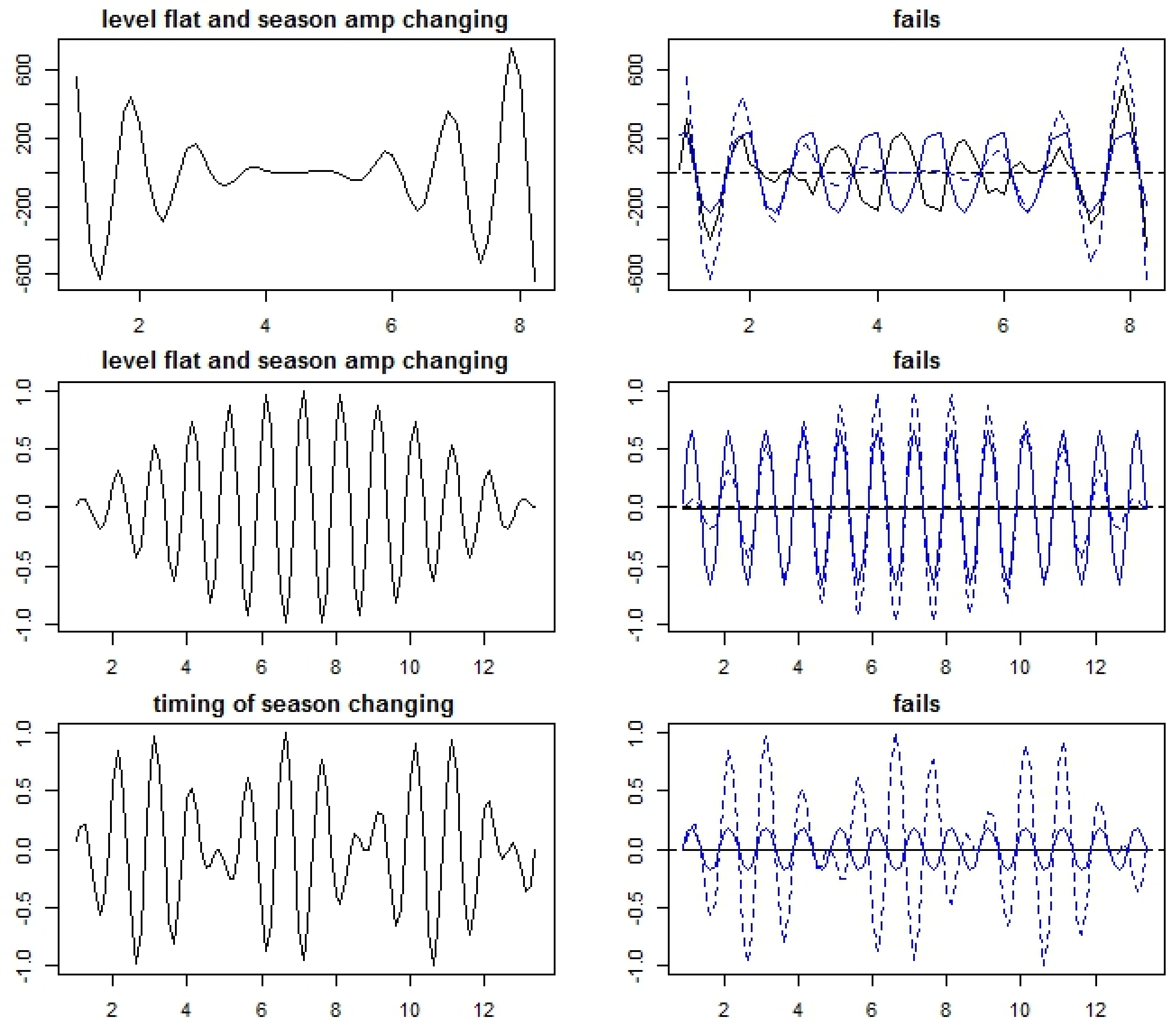
Log qtrly Kerala catch

Line that goes through the data. Like a time-varying intercept/level

Time-varying seasonal component.

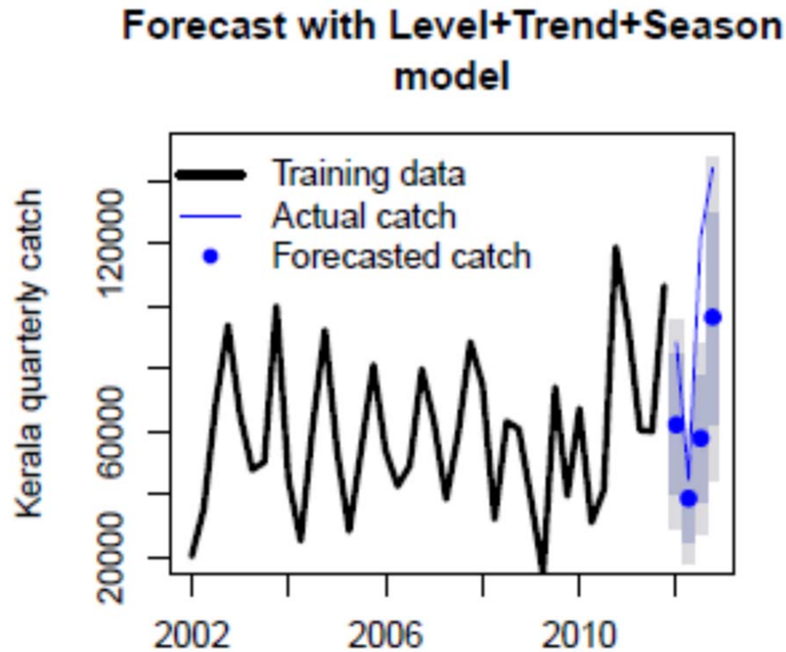



season and changing level
works
season and changing level
works
level flat and season amp changing
works



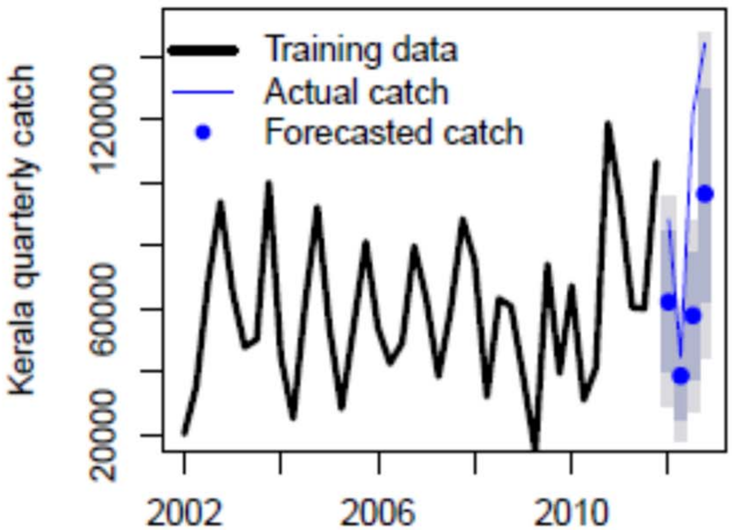
level flat and season amp changing
fails
level flat and season amp changing
fails
timing of season changing
fails

Exponential smoothing model forecast

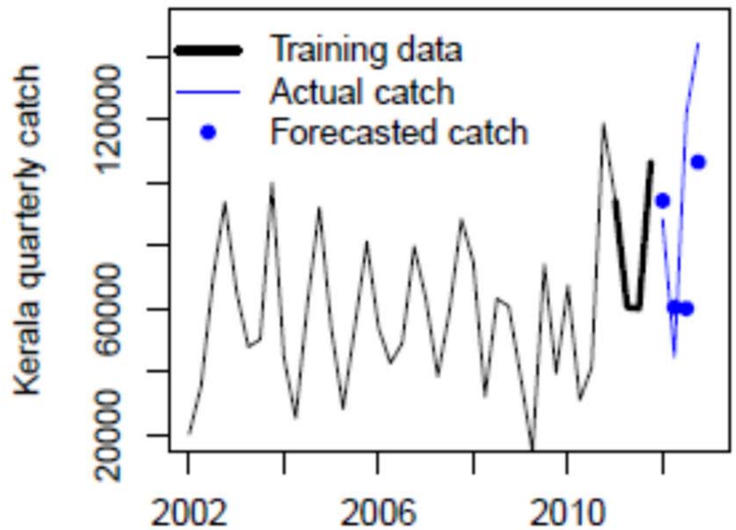


One step ahead forecast tests (out of sample)

Forecast with Level+Trend+Season model



Forecast using last year



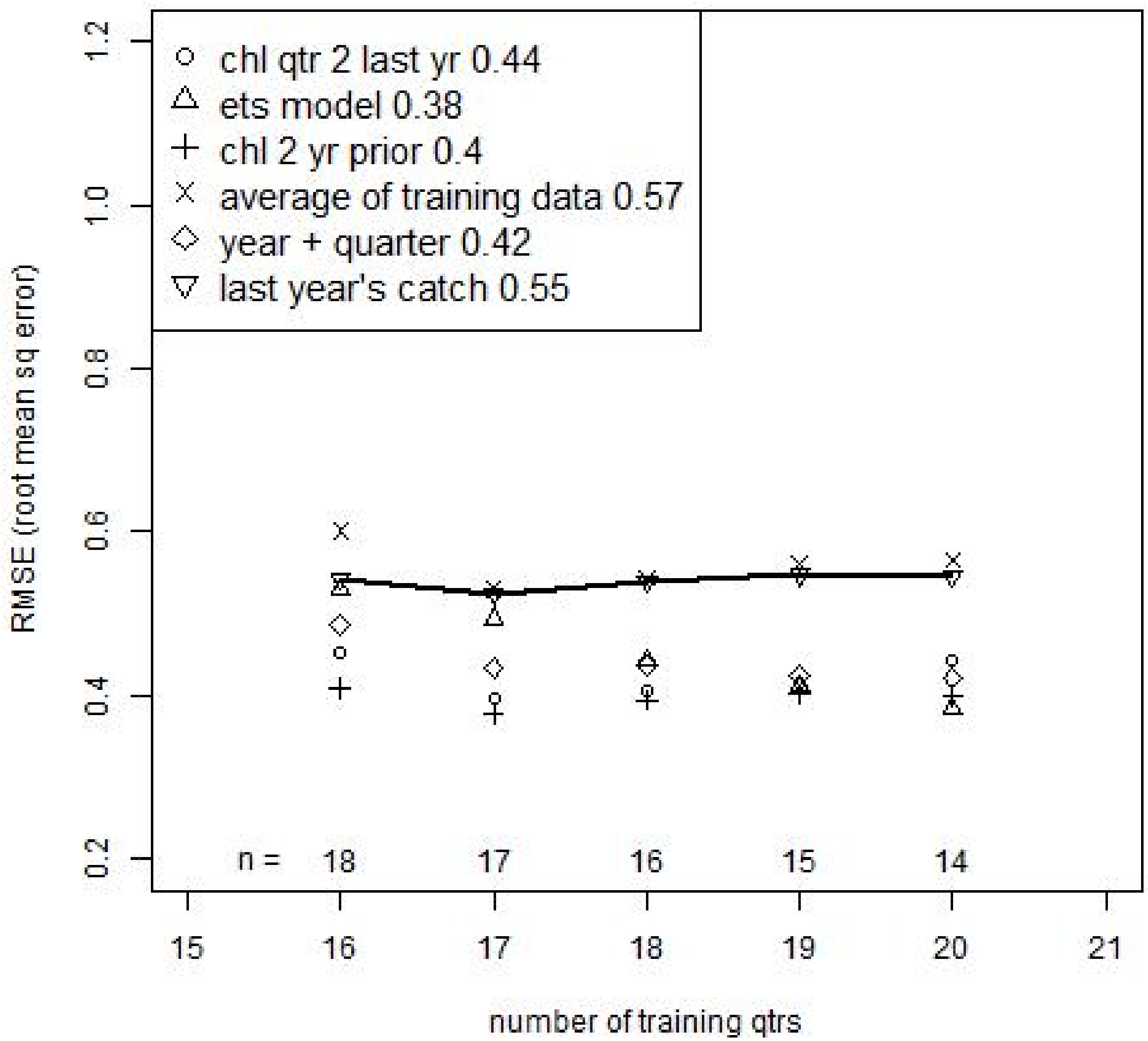
- Predict one qtr ahead for a series of training windows of different sizes
- Store the error (predict – observed)
- Compute the Root mean square error (RSME) for prediction performance

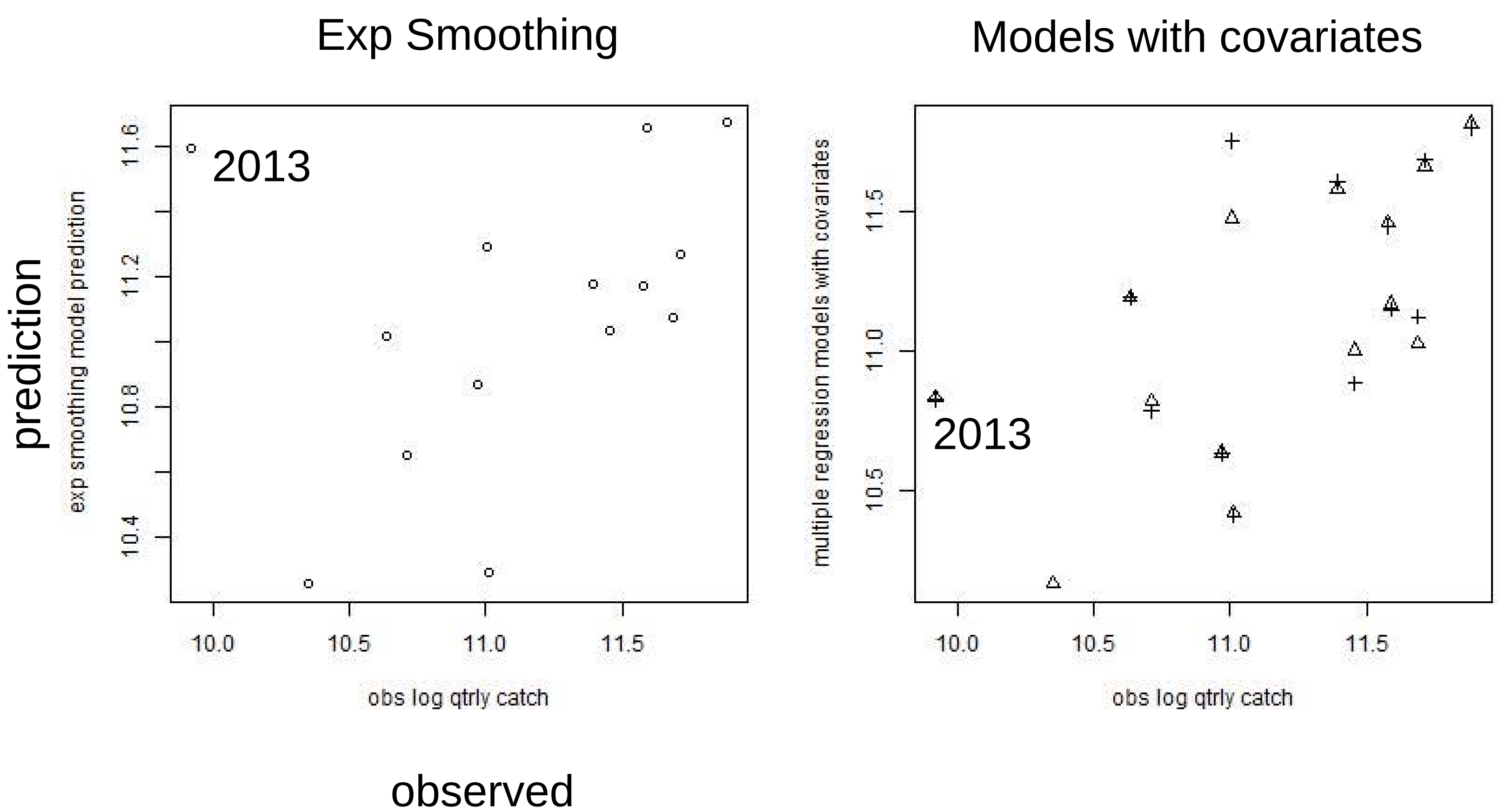
NOAA FISHERIES

Models I will test

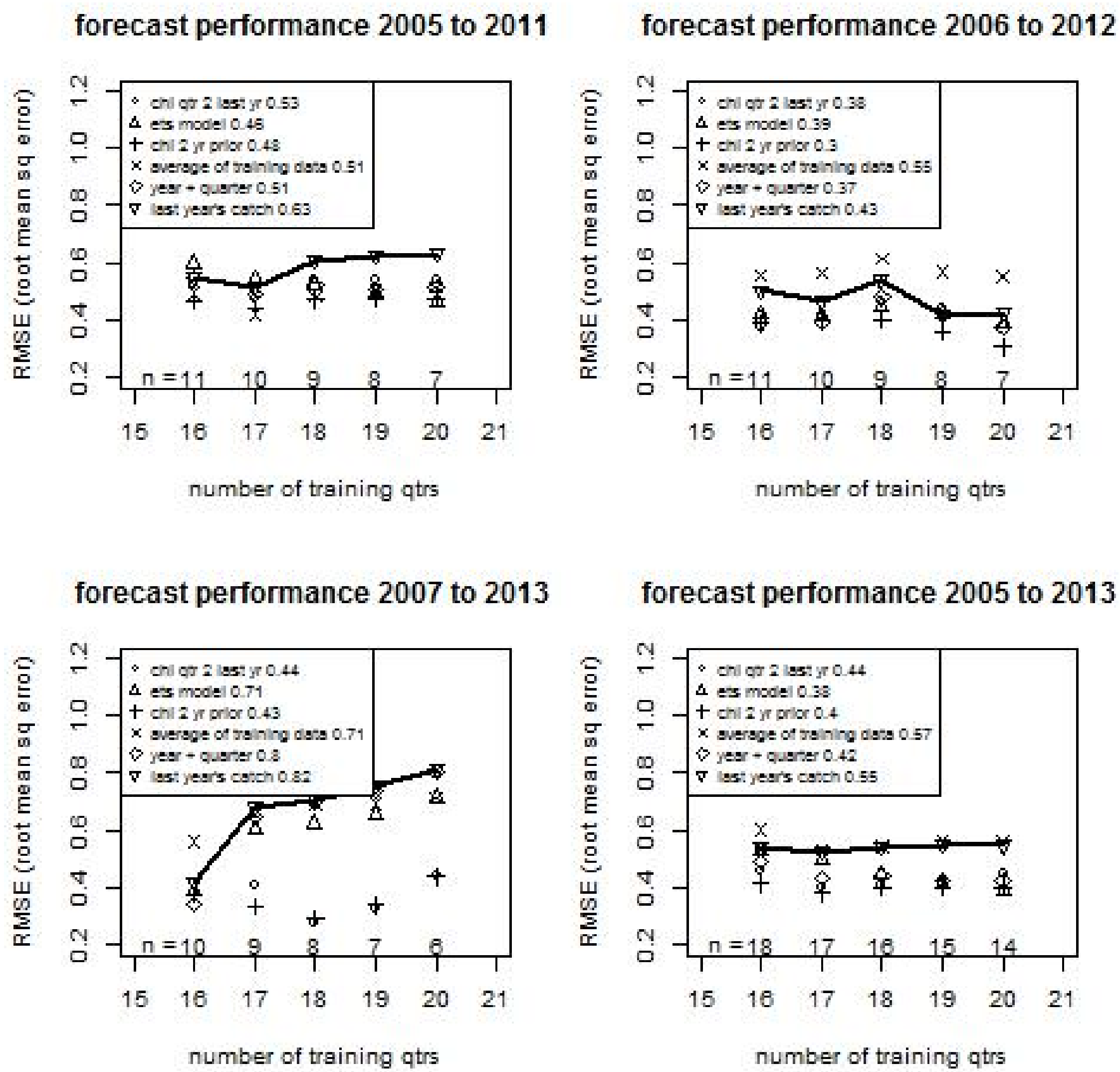
- Linear regression with quarter factor (season), chlorophyll anomaly from quarter 2 (pre-monsoon) of the previous year
- Linear regression with quarter factor (season), chlorophyll anomaly from 8 quarters (2 years) prior
- Exponential smoothing model with time-varying quarter factors and time-varying level
- Linear regression with year effect and quarter
- Linear regression intercept only (flat level model)
- Use last year’s catch from the same quarter as the predictor

forecast performance 2005 to 2013




Exp Smoothing
Models with covariates
prediction
exp smoothing model prediction
multiple regression models with covariates
2013
2013
10.4
10.8
11.2
11.6
10.5
11.0
11.5
10.0
10.5
11.0
11.5
10.0
10.5
11.0
11.5
obs log qtrly catch
obs log qtrly catch
observed

Same take-home message with slightly different testing year ranges. Exp smoothing and covariate models performing similarly



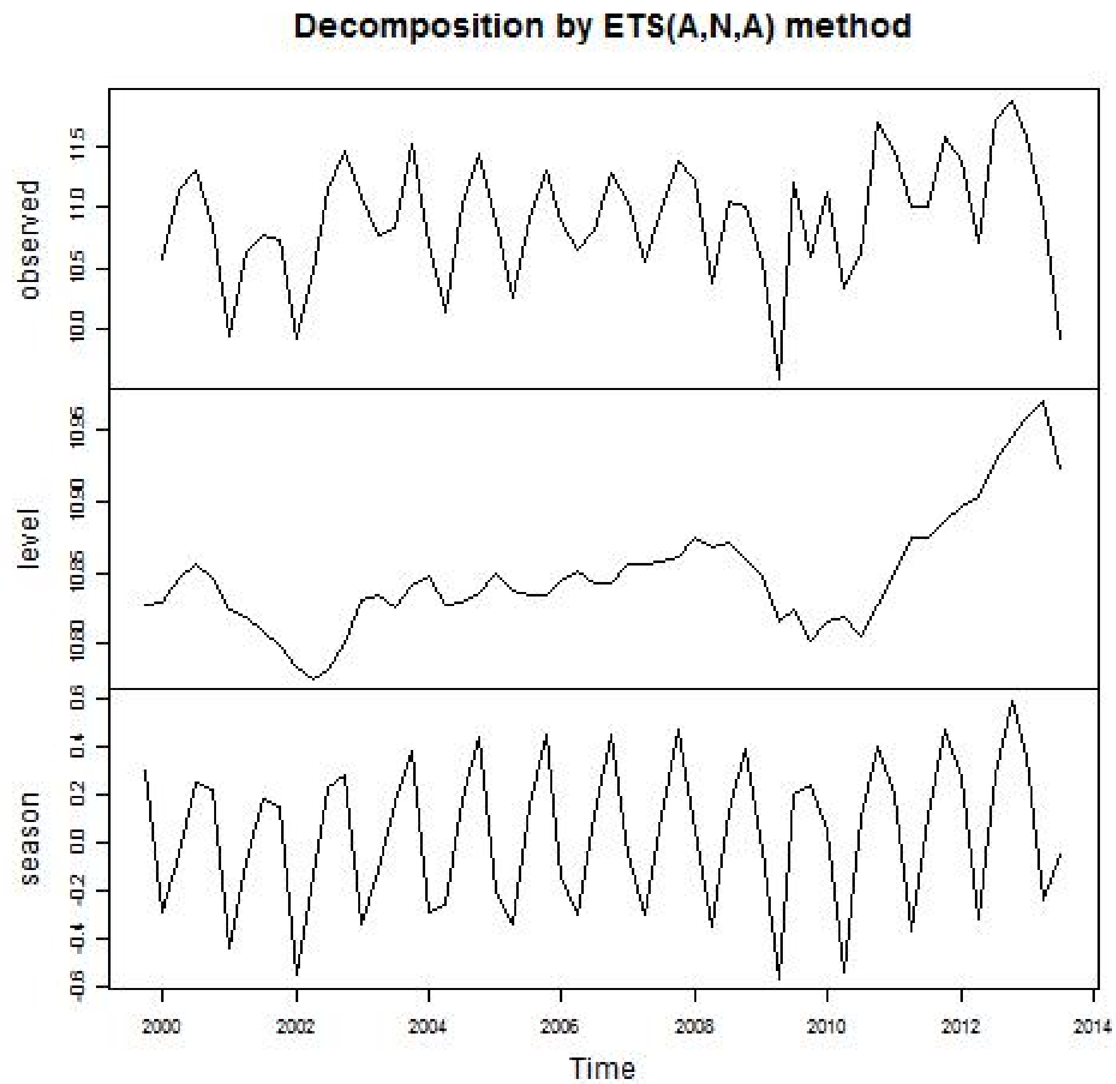
Thoughts and plans for the forecast model development

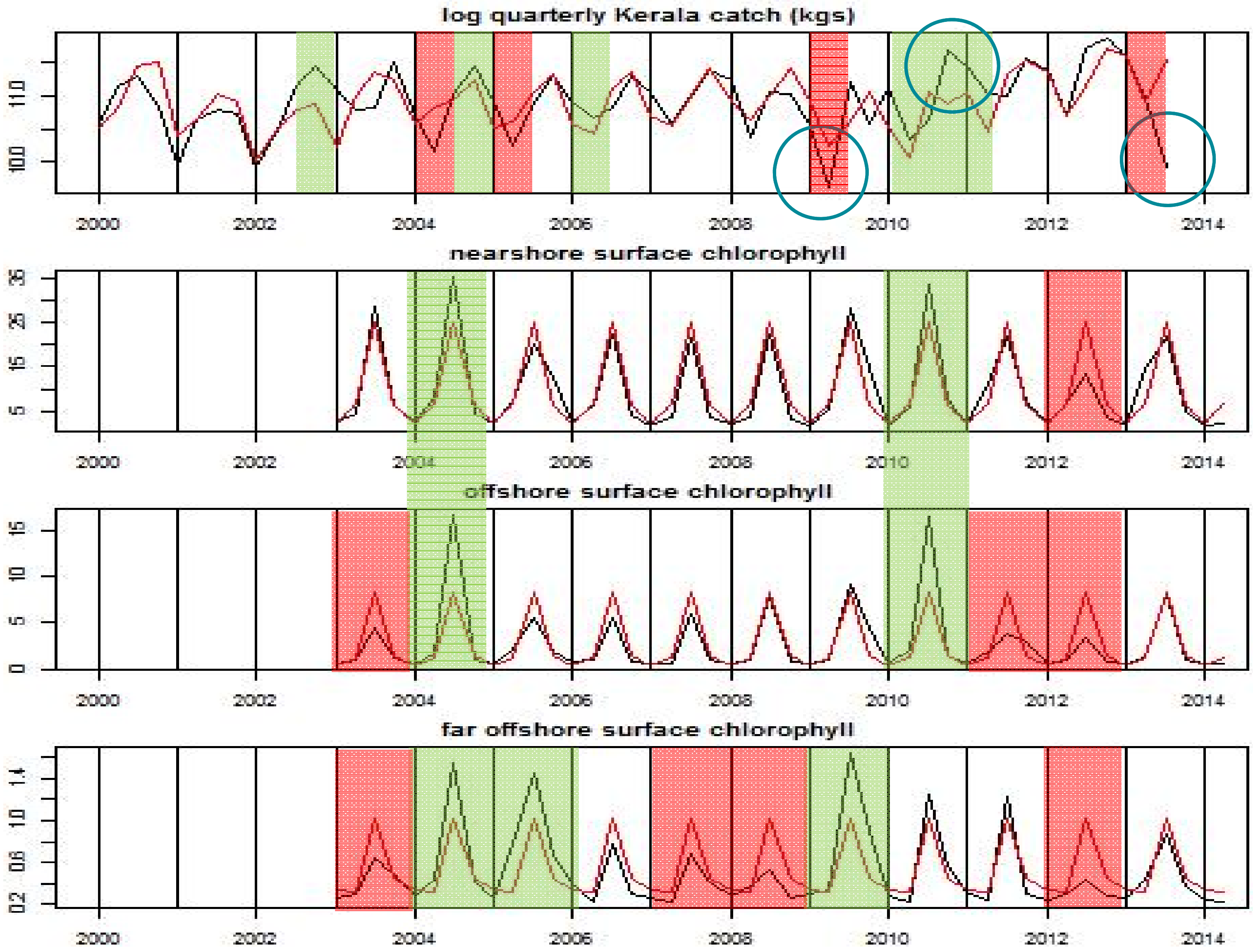
- Need more data for testing and building models
- My forecast models seem to be an improvement over season model and last year's catch model.
- But….I used catch data up to the previous qtr. It takes 12 months to process the data. So in reality you only catch data from 12 months prior not the previous qtr.
  - Probably will kill the exponential smoothing model predictions
- Work on using covariates (SSH) as the 'season' metric rather than estimating using factors.
- Test ARMA models
  - Doubtful unless I can use much more training data than 16 qtrs.

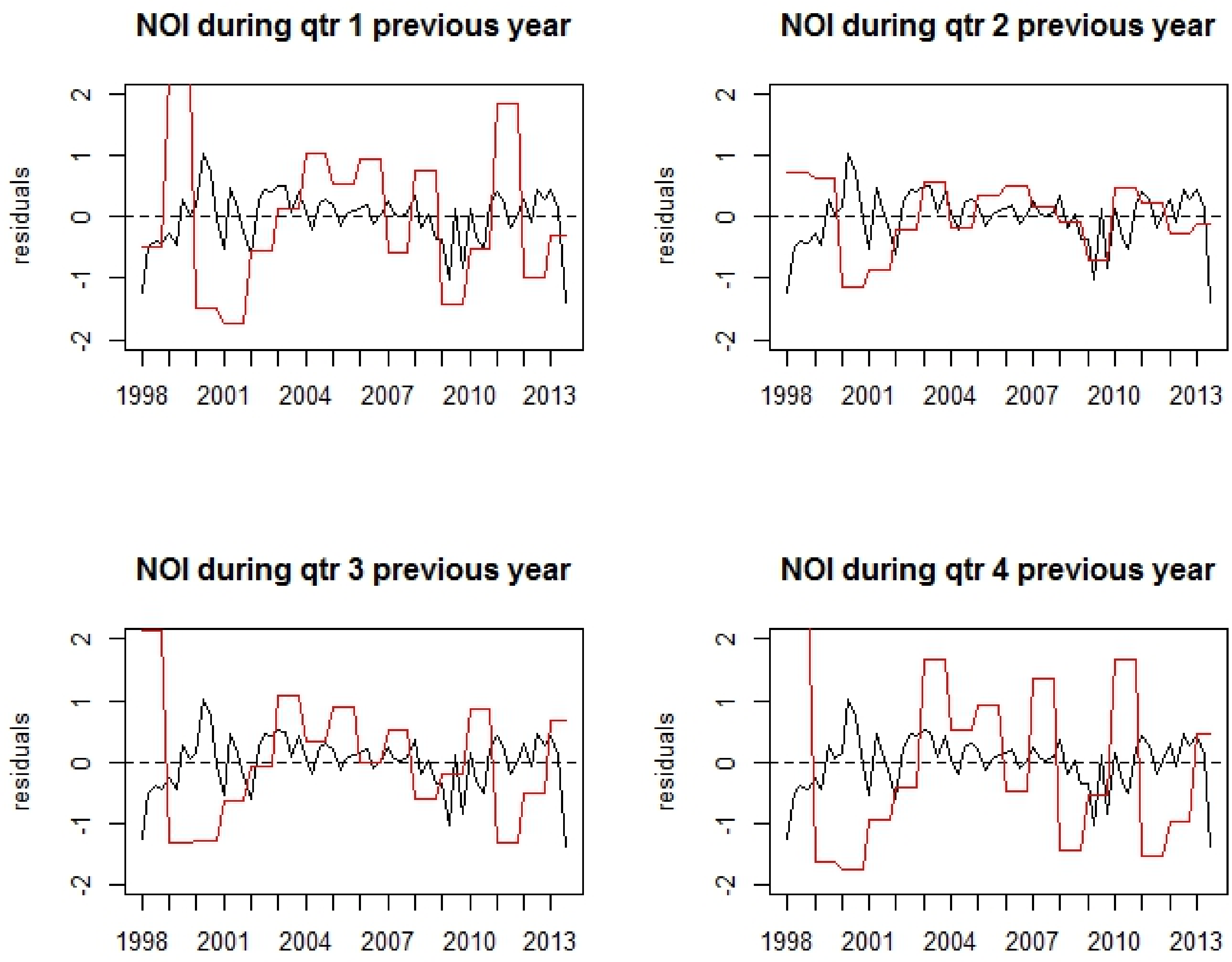
Log qtrly Kerala catch

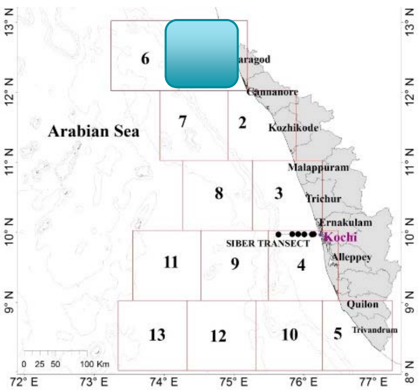
Line that goes through the data. Like a time-varying intercept/level

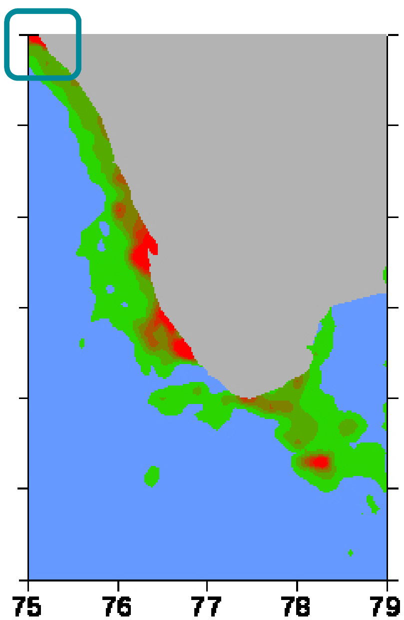
Time-varying seasonal component.

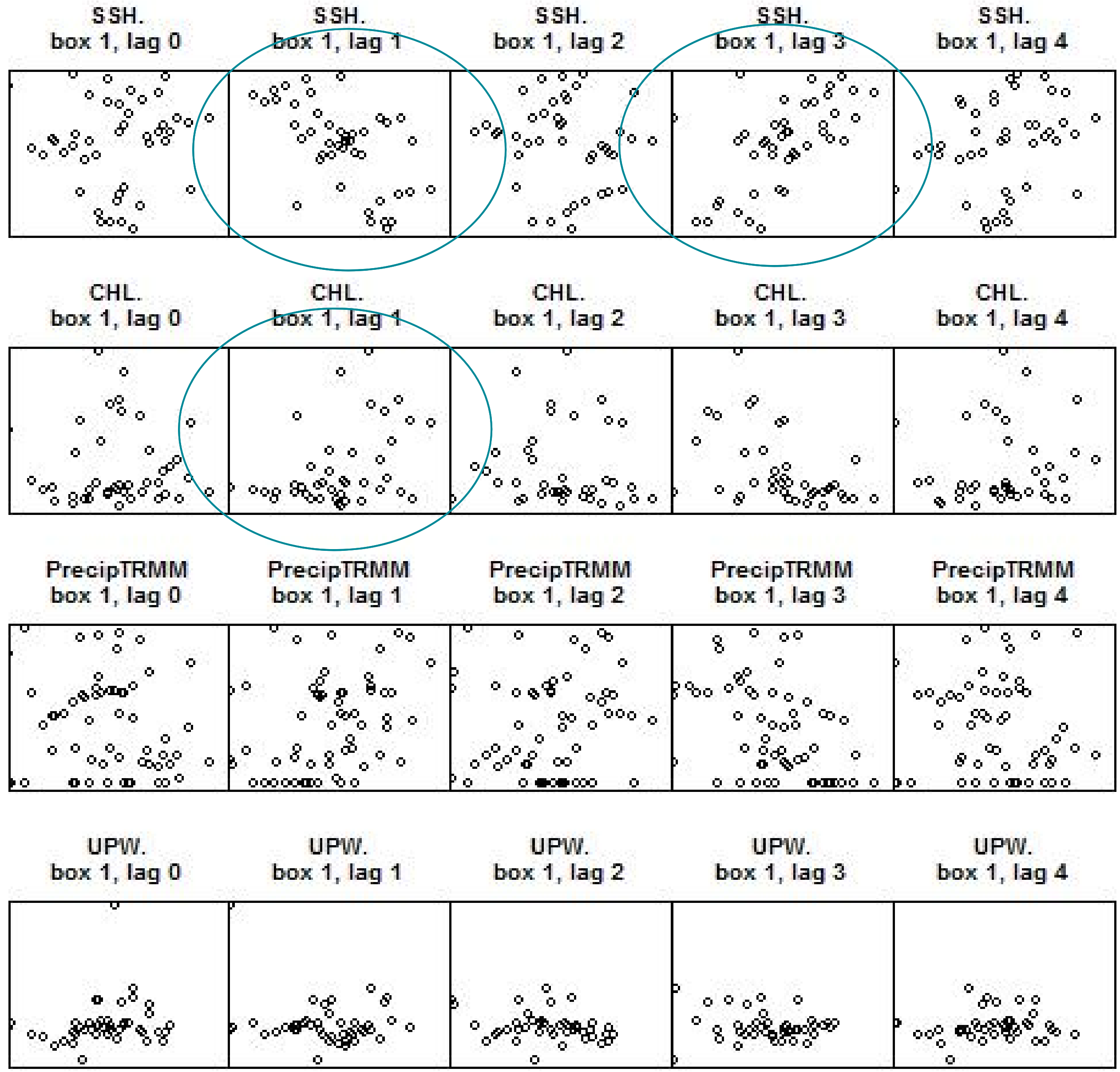



log quarterly Kerala catch (kgs)
nearshore surface chlorophyll
offshore surface chlorophyll
far offshore surface chlorophyll
2000
2002
2004
2006
2008
2010
2012
2014


NOI during qtr 1 previous year
NOI during qtr 2 previous year
NOI during qtr 3 previous year
NOI during qtr 4 previous year
residuals
2
1
0
-1
-2
1998
2001
2004
2007
2010
2013



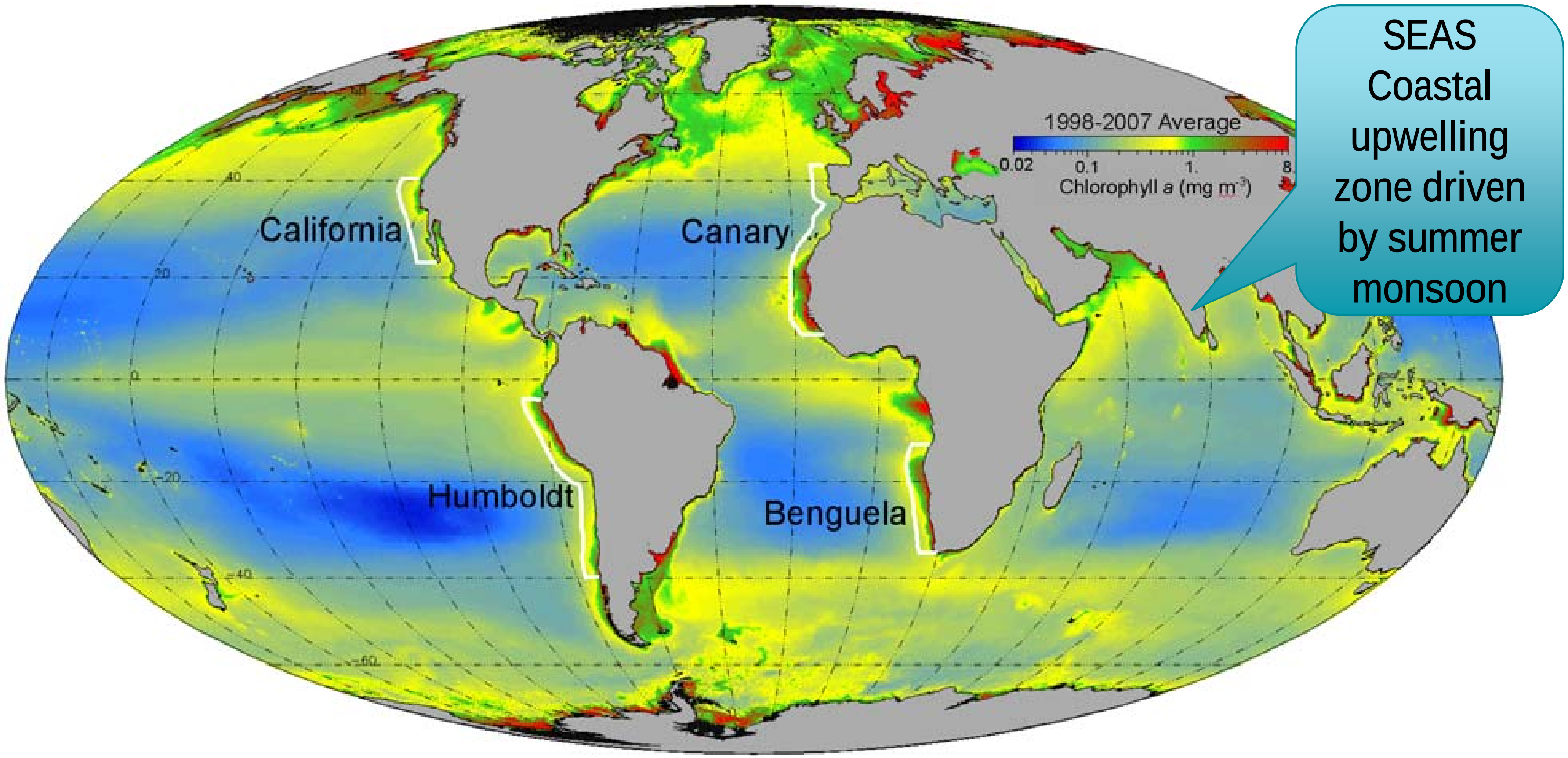




SST left off for space. Less correlated than SSH, more than Chl-a

The South east Arabian Sea is one of world's major upwelling zones and one of the most productive regions of the world's oceans

Has a strong seasonal upwelling system driven by winds during the monsoon season (May-Sept)



Surface Chl-a 1998-2007 Average