Center for the Advancement of Population Assessment Methodology

Comisión Interamericana del Atún Tropical
Inter-American Tropical Tuna Commission

CAPAM

CIAT IATTC

# Model Weighting using cross validation and hindcasting

## Virtual meeting, 28 Nov – 2 Dec (8am to 11am - San Diego)

# Background papers

- Kell, L. T., Kimoto, A., and Kitakado, T., 2016. Evaluation of the prediction skill of stock assessment using hindcasting. Fisheries Research, 183: 119–127.
- Kell, L.T., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., and Fu, F. 2021. Validation of stock assessment methods: is it me or my model talking? ICES Journal of Marine Science, 78(6), 2244–2255.

# Concept

- Evaluate the model based on how well it can predict out of sample data
- A portion of the data to train the model and a portion of the model to test the model
- Works if data weighting or random effects are not correct
- Can inform whether there is overfitting or bias
- Can’t validate situations not observed

# Issue 1: Time series

- Stock assessment based on a dynamic model so autocorrelation is inherent
- Simple cross validation does not work for time series models
- Particularly if you are using it for management advice in the following year
- Use one-step-ahead cross validation (hindcasting)

# Issue 2: We don’t observe the quantities of interest

- Quantities of interest
  - F, B, $F/F_{MSY}$, $B/B_{MSY}$
- Observations
  - Catch, Index of abundance (CPUE), age or length compositions, conditional age-at-length, tagging, …
- Predict the observed data -> good model -> good estimates of management quantities

# Example: Predicting catch

- Yellowfin tuna in the EPO
- Change from effort regulation to catch regulation

- Question: What is the appropriate annual catch limit?
- How: Given $F_{MSY}$ what is the next years catch
- One step ahead test: Given the observed effort level, can the model predict the next years catch?
- Result: within 50% to 200% of the actual catch
- Implication: Need an index of recruitment (unless it is catchability)

- Follow-up: Development of a weekly depletion estimator for in-season management (never used)

# Decision 1: What data to predict

- Data types
  - Index of relative abundance
  - Age/length composition
  - Mean age/length
  - Conditional age-at-length
  - Tagging
  - Other
- Rationale
  - Closest to management quantity
  - Sensitive to model misspecification

- Recommended
  - Most reliable index of abundance related to spawning biomass
  - Mean age/length of the index
  - Mean age/length of a "recruitment" fishery if recruitment is important to the management quantities

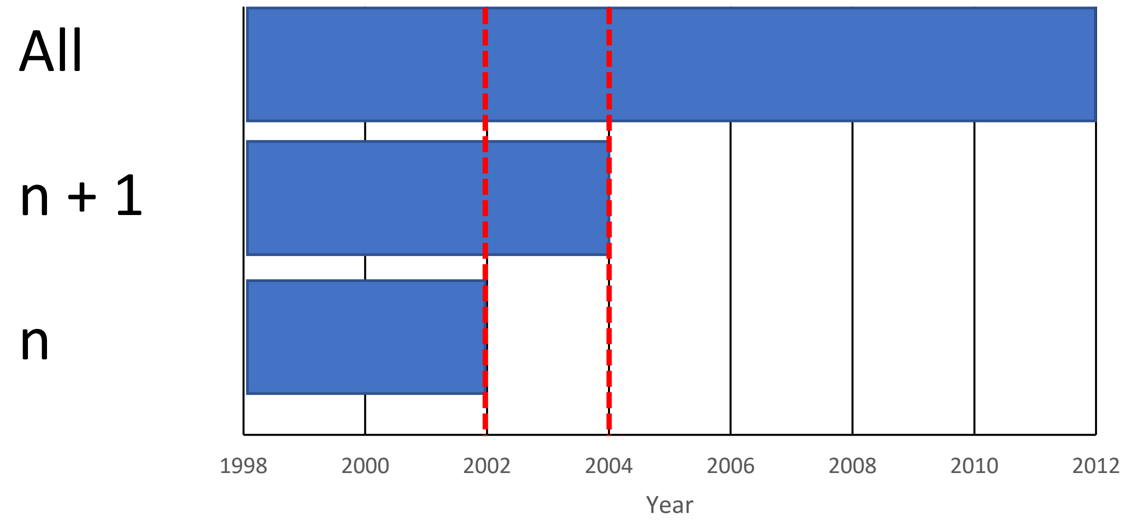
# Decision 2: What data to remove

- Series
- Data type
- Fleet
- Time blocks
- Individual points
- Combinations of series or data types
- Allows for data conflicts to be evaluated

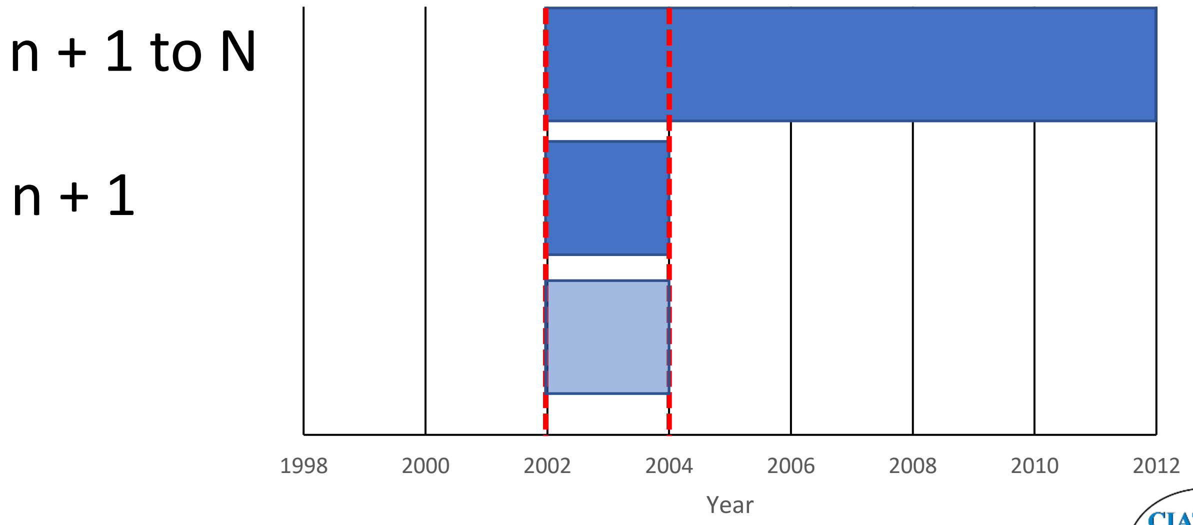
- Recommended
  - One year of data in the previous slide

# Decision 3: What data to include and what to predict

## Other data to include (e.g. length comps)




## Data to predict



# Decision 4: How many years to remove

- Depends on reason for using hindcasting
- 1 year
- Management cycle (e.g. assessment every 3 years)

- Recommended
  - One year

# Decision 5: Prediction measures

- Root mean squared error (RMSE)
  - Sensitive to outliers
- Correlation
- Mohn's rho
  - Used for retrospective analysis
- Relative error
- Mean absolute scaled error (MASE)
  - Compared to naive prediction (last years value)
  - Scale invariant, symmetry, interpretability, asymptotic normality
- Likelihood
  - Convert to probabilities?
  - Need good estimate of variance parameter

- Recommended
  - Likelihood

# Mean absolute scaled error (MASE)

Model prediction:
Based on data from
previous year
lag h

For a peel of $n$ and a horizon of $h$ years

$$MASE = \frac{\frac{1}{n+1}\sum_{t=T-n}^{T}\left|y_t - \hat{y}_{t|t-h}\right|}{\frac{1}{n+1+h}\sum_{t=T-n-h}^{T}\left|y_t - y_{t-h}\right|}. \qquad (5)$$

Simple prediction:
Equal to last observed
value (previous year
lag h)

Kell, L.T., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., and Fu, F. 2021. Validation of stock assessment methods: is it me or my model talking? ICES Journal of Marine Science, 78(6), 2244–2255.

CIAT

IATTC