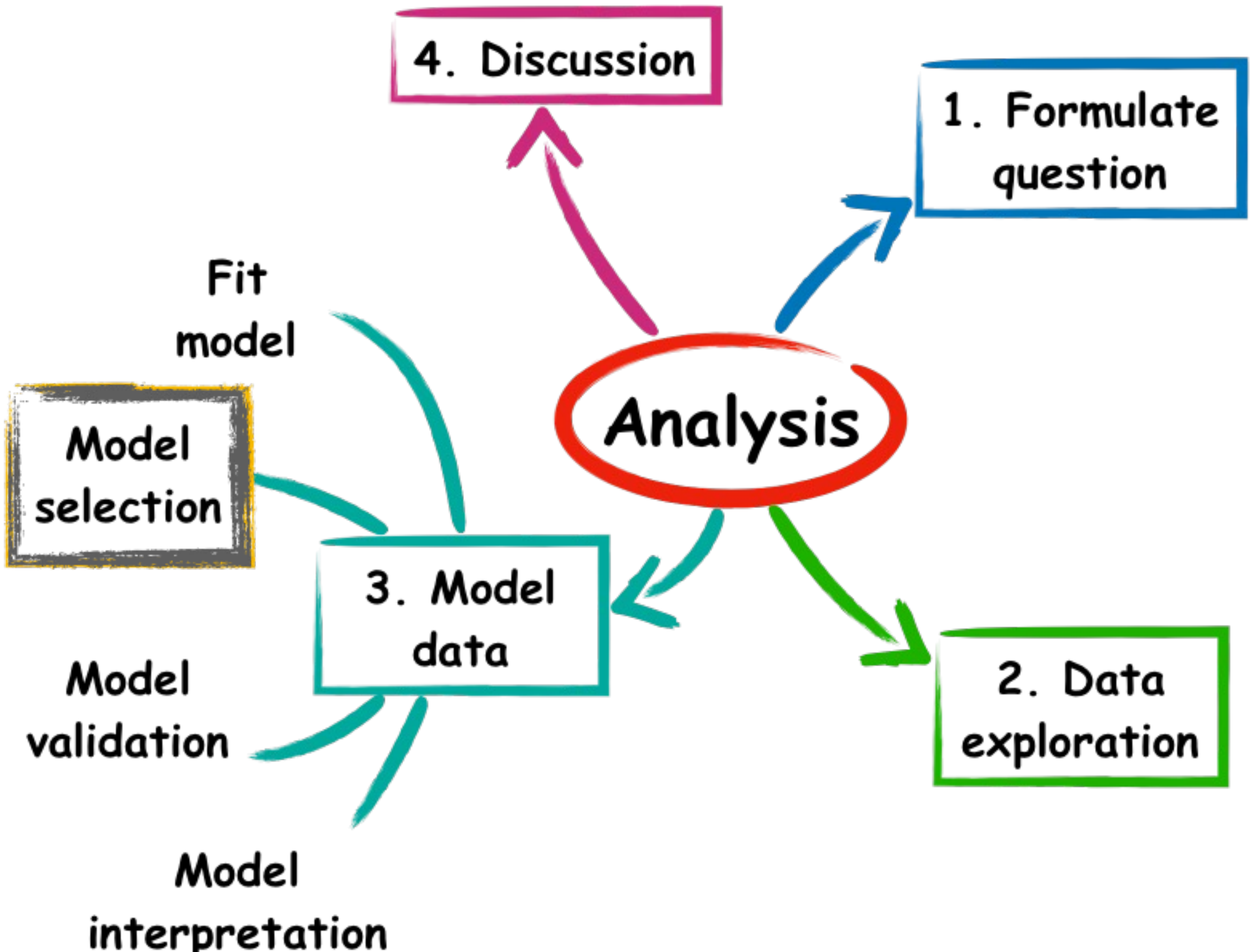
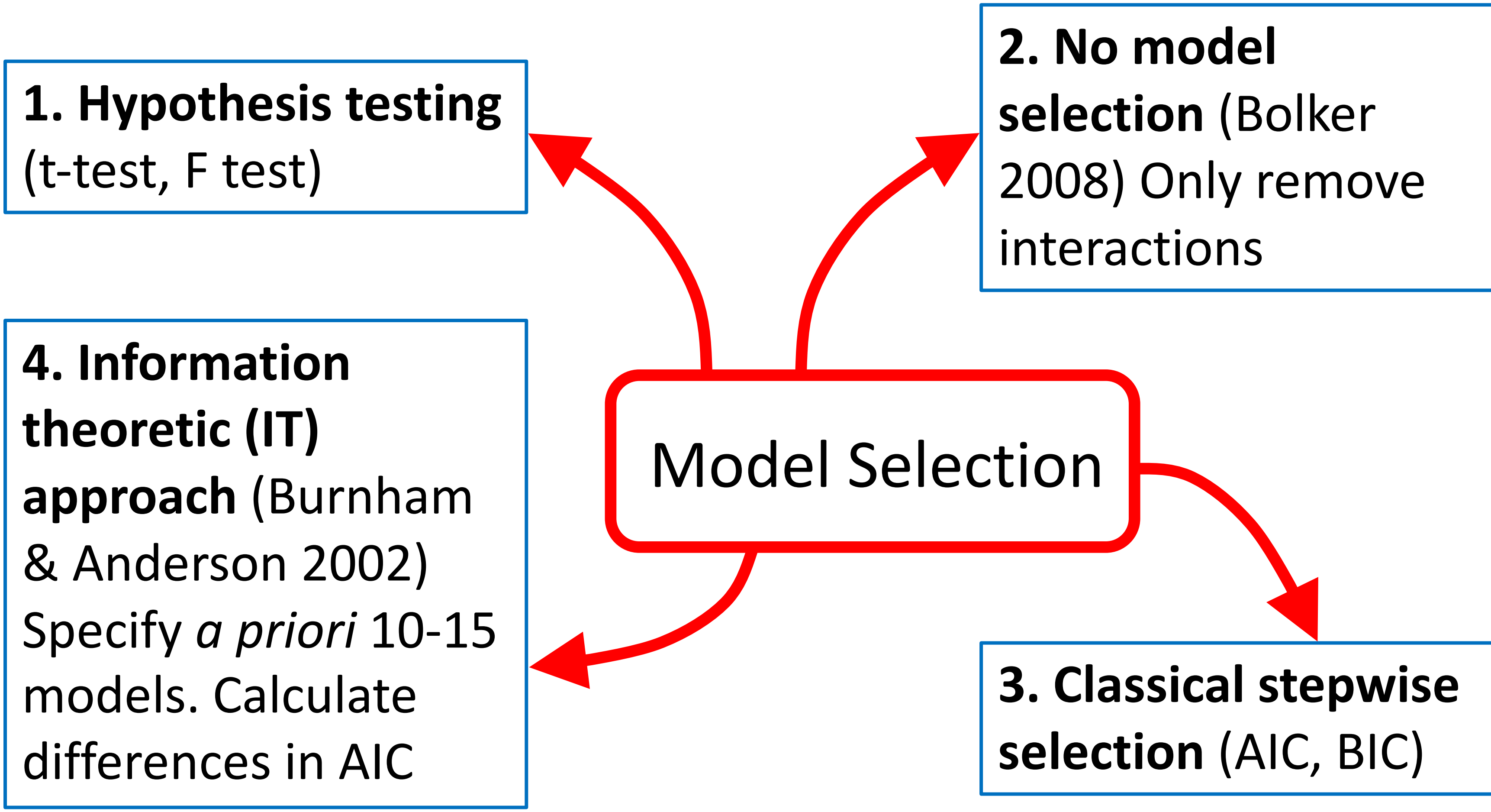

4. Discussion
1. Formulate question
Analysis
Fit model
Model selection
3. Model data
2. Data exploration
Model validation
Model interpretation

Model selection

1. Controversial area of statistics
2. Several alternatives – different "schools of thought"
3. Depends on your aim in fitting a model
4. ...and your study system


Model Selection
1. Hypothesis testing (t-test, F test)
2. No model selection (Bolker 2008) Only remove interactions
3. Classical stepwise selection (AIC, BIC)
4. Information theoretic (IT) approach (Burnham & Anderson 2002) Specify *a priori* 10-15 models. Calculate differences in AIC

1. Hypothesis testing

- Drop least significant term
- Refit model
- Continue until only significant terms

I suggest never using this approach

- The best-fitting model may include non-significant terms
- Referees will (rightly) criticize this approach
- Consider what a P-value actually represents

2. Do nothing

- Perfectly valid (and you can’t be criticized for the model selection approach you might otherwise use)
- Illustrates which terms in the model have significance and which don‘t (this could be your main question)
- *A priori* you selected certain covariates, so why remove them?
- (do remove collinear terms)

3. Classical stepwise selection

- Use backward (start with full model and remove terms) or forward (start just with intercept and add terms) selection
- Use Akaike Information Criteria (AIC) to arrive at best-fitting model (also BIC, and for Bayesian models DIC, WAIC)

4. Information theoretic (IT) approach

- Formulate (*a priori*) 10-15 alternative models
- Run all models, then compare using AIC
- Advocated by respected statisticians (Burnham & Anderson, 2002)
- A very powerful approach
- ....but requires a lot of information/understanding
- Usually the case in fisheries models

model	fitted model	source
M01	temperature + salinity	Heuts (1947)
M02	presence/absence of fish predators	Hoogland *et al*. (1956)
M03	latitude x longitude	Münzig (1963)
M04	temperature	Wootton (1976)
M05	presence/absence dragonfly larvae	Reimchen (1994)
M06	pH	Giles (1983)
M07	elevation	Raeymaekers *et al*. (2007)
M08	salinity	Myhre & Klepaker (2009)
M09	presence/absence *Schistocephalus solidus*	Morozińska-Gogol (2011)
M10	presence/absence *Pungitius pungitius*	MacColl *et al*. (2013)
M11	turbidity + presence/absence of fish predators	Reimchen *et al*. (2013)
M12	pH + presence/absence of fish predators	Spence *et al*. (2013)
M13	pH + presence/absence of fish predators + turbidity	Klepaker *et al*. (2016)
M14	presence/absence of fish predators + *P. pungitius*	Magalhaes *et al*. (2016)
M15	temperature + standard length + pH	this study

Run all 15 models and compare with AIC

My suggestion

1. Use IT when possible
2. Alternatively, depending on the aims of your study, either
   - Perform no selection, or
   - Manual backward selection
3. Avoid using hypothesis testing

How to deal with zero catches?

- **Do not ignore zeros** - these are critical data!
- Use an appropriate distribution that can accommodate zero observations
- Simulate from your model to ensure the model accommodates the proportion of zeros in the data
- We will do this (Hilsha analysis)

How many zeros is too many?

- No specific threshold
- Fit model, then simulate from it
- Does the observed number exceed the predicted (by a lot)

What distribution is appropriate for (many) zeros?

- Gaussian (?), Poisson, negative binomial, Bernoulli, binomial
- Model validation: check by simulating from model and compare proportion of zeros in simulated data sets with observed proportion – they should match
- Use ‘testZeroInflation’ command in ‘DHARMa’ package
- We will do this (Hilsha analysis)

Why do we get lots of zeros?

- Unsuitable conditions – no catch
- Suitable conditions – no catch
- Suitable conditions – not catchable
- Suitable conditions – make error

What type
of zeros
do you
have?

How to handle lots of zeros

- Fit zero-inflated (mixture) models
- Fit zero-adjusted (hurdle) models

ZIP, ZAP!

- Zero-inflated models differ from zero-adjusted models
- Zero-inflated models - model zeros as counts (some of which are zero)
- Zero-adjusted models explicitly model zeros as a Bernoulli model, and counts (zero-truncated data) using Poisson, NB, Gamma

Zero-inflated models

- Model data in two parts:
  - Binomial part; zeros vs. count (use binomial distribution)
  - Zero-truncated data, using Poisson, negative binomial, gamma
- Able to identify which variables result in a catch (binomial part) and if a catch occurs, the size of the catch (zero-truncated part)
- We will use a ZINB model with the Hilsha analysis

Tweedie distribution

- A family of distributions
- Not widely used
- Easy to implement with the ‘glmmTMB’ package
- Able to generate a compound Poisson-Gamma distribution

Approach

1. Formulate the question

Standardise CPUE for Bangladesh hilsha catch