

# “Maintain or Improve” How do we judge that?

Graham McBride

Principal Scientist – Water Quality

NIWA, Hamilton

NZFSS Conference, 21 November 2017



# Part 1: Improve?

# What's the problem?

- Trend analysis for ~500 sites had 38% - 74% trends 'uncertain', because  $P$ -value  $> 0.05$  (Larned *et al.* 2016)

# What's the problem?

- Trend analysis for ~500 sites had 38% - 74% trends 'uncertain', because  $P$ -value  $> 0.05$  (Larned *et al.* 2016)
- Loss of potential information

# What's the problem?

- Trend analysis for ~500 sites had 38% - 74% trends 'uncertain', because  $P$ -value  $> 0.05$  (Larned *et al.* 2016)
- Loss of potential information
- Something deficient in traditional trend tests?
  - Answer: **Yes!**
  - Tests a dumb hypothesis: 'Is the true trend slope exactly zero?'
  - Answer: **No!** Hypothesis is *a priori* false
  - Everything changes over time

# What's the problem?

- 'Trend' undefined
  - It's a gloss put onto the test result
- $P$ -values for such hypotheses do not serve as a weight-of-evidence metric
  - Get ever-smaller with sampling density
- *Statistical* significance is not *environmental* significance
- 'Not statistically significant' does **not** mean 'nothing to be concerned about'

# What's the solution?

- Define 'trend' as a non-zero slope
- So there is *always a trend*, however small
- Abandon hypothesis/significance tests!
- Instead calculate the probability that the trend slope is +ve (or -ve)
- Replace Type I or Type II errors by two misclassification probabilities
  - Inferring a positive trend when the true trend is negative, and *vice versa*
- Use these probabilities in a 'likelihood' categories table

# 'Likelihood' categories

---

|                        |         |
|------------------------|---------|
| Virtually certain      | 99–100% |
| Extremely likely       | 95–99%  |
| Very likely            | 90–95%  |
| Likely                 | 67–90%  |
| About as likely as not | 33–67%  |
| Unlikely               | 10–33%  |
| Very unlikely          | 5–10%   |
| Extremely unlikely     | 1–5%    |
| Exceptionally unlikely | 0–1%    |

---

# It's Bayesian, a different probability concept

- 'Probability' has more than one definition

# It's Bayesian, a different probability concept

- 'Probability' has more than one definition
- Especially, frequentist versus Bayesian

# It's Bayesian, a different probability concept

- 'Probability' has more than one definition
- Especially, frequentist versus Bayesian
- Most interpretations of confidence intervals are (unwittingly) Bayesian
  - 'I am 95% sure that the true mean lies in this interval'
  - Invoking 'non-informative' prior distributions
  - That's **not** frequentist

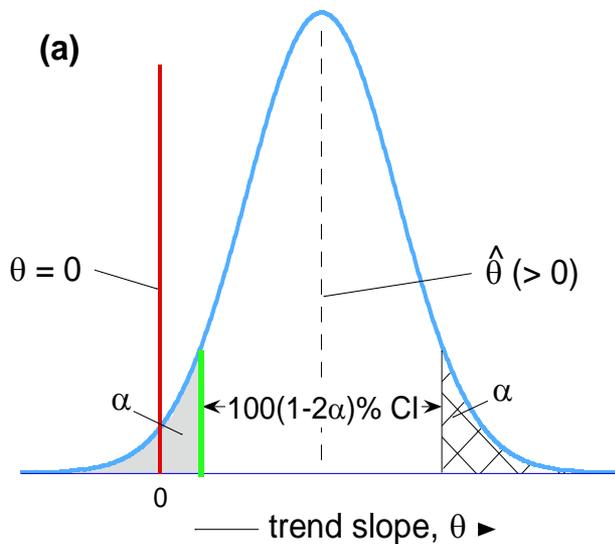
# It's Bayesian, a different probability concept

- 'Probability' has more than one definition
- Especially, frequentist versus Bayesian
- Most interpretations of confidence intervals are (unwittingly) Bayesian
  - 'I am 95% sure that the true mean lies in this interval'
  - Invoking 'non-informative' prior distributions
  - That's **not** frequentist
- We explicitly adopt Bayesian probability

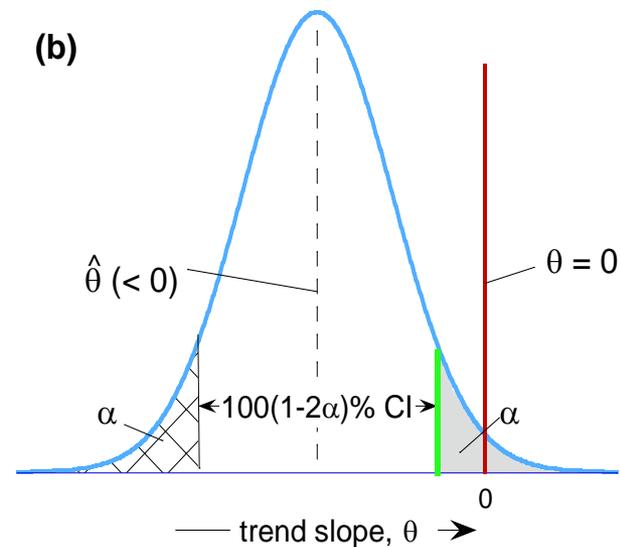
# It's Bayesian, a different probability concept

- 'Probability' has more than one definition
- Especially, frequentist versus Bayesian
- Most interpretations of confidence intervals are (unwittingly) Bayesian
  - 'I am 95% sure that the true mean lies in this interval'
  - Invoking 'non-informative' prior distributions
  - That's **not** frequentist
- We explicitly adopt Bayesian probability
- Calculation straightforward, already embedded in *Time Trends*  
<http://www.jowettconsulting.co.nz/home/time-1>

# In a diagram



Slope is inferred positive if the lower CI limit (green line at right edge of left shaded region) is greater than zero (red line), as shown.



Slope is inferred negative if the upper CI limit (green line at left edge of right shaded region) is greater than zero (red line), as shown.

# Using Larned data: Probability of an improving trend

| <b>Category</b>           | <b>DRP</b> | <b>ECOLI</b> | <b>NH4N</b> | <b>NO3N</b> | <b>TP</b> | <b>TN</b> | <b>CLAR</b> |
|---------------------------|------------|--------------|-------------|-------------|-----------|-----------|-------------|
| Virtually certain         | 116        | 33           | 60          | 71          | 178       | 49        | 60          |
| Extremely likely          | 42         | 35           | 18          | 39          | 54        | 19        | 50          |
| Very likely               | 31         | 28           | 17          | 28          | 31        | 9         | 26          |
| Likely                    | 63         | 69           | 17          | 63          | 63        | 42        | 64          |
| About as likely as<br>not | 48         | 92           | 22          | 73          | 38        | 29        | 67          |
| Unlikely                  | 28         | 77           | 19          | 59          | 32        | 28        | 60          |
| Very unlikely             | 16         | 19           | 11          | 31          | 8         | 14        | 14          |
| Extremely unlikely        | 8          | 27           | 20          | 33          | 8         | 12        | 26          |
| Exceptionally<br>unlikely | 39         | 16           | 22          | 114         | 9         | 41        | 19          |

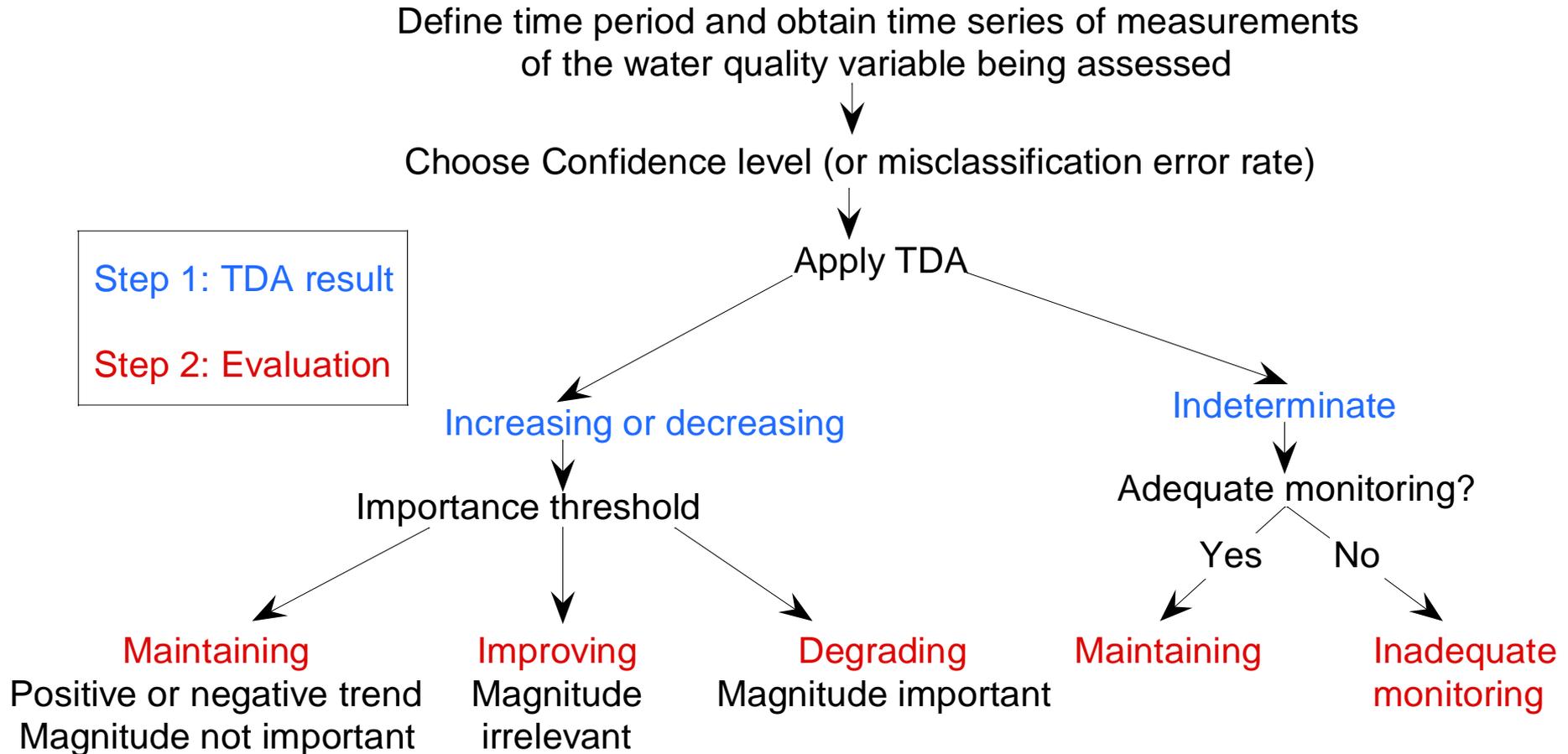
# Using Larned data: Probability of an improving trend

- No results put into a 'not significant' box
- Gives information about every result
- Percentage of sites for which *E. coli* trend is other than 'about as likely as not' falls from 74% to 23%.
- Up to the user to decide probability cutoff

## Part 2: Maintain?

- Under development
- Extra difficulty
  - What trend magnitude would be considered as environmentally significant?
  - Over what time period?
  - Must state these
- Introduces subjectivity
  - Can't be avoided

# Part 2: Improve? (under development)



# Conclusions

- Abandon trend tests and associated  $P$ -values
- Use TDA (Trend Direction Assessment)
- To address “Improve”, enter TDA results into likelihood category table
- To address “maintain” must define environmentally significant trend magnitude, and for what period
- Beware of censored values