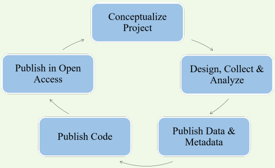
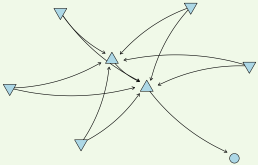
Use of open science to inform restoration projects in estuaries: A Tampa Bay example

Marcus W. Beck[1], Ed Sherwood, Kirsten Dorans, Jessica Renee Henkel, Kathryn Ireland, Patricia Varela

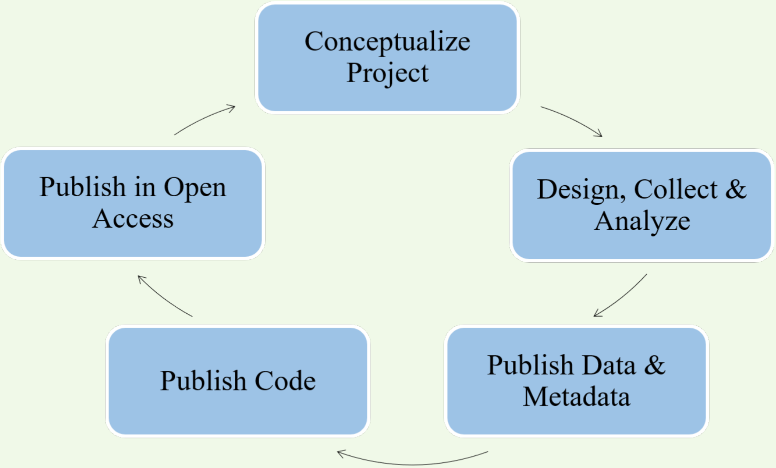
[1]Southern California Coastal Water Research Project, Costa Mesa, CA
marcusb@sccwrp.org, Phone: 714-755-3217

April 23, 2018





Open science workflow

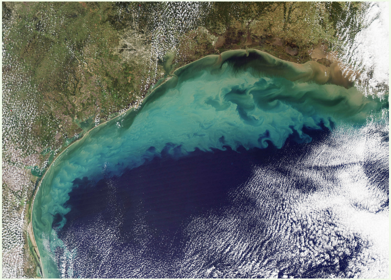


Modified from Hampton et al. 2015. The Tao of open science for ecology. Ecosphere 6(7):1-13.

Open Science for Synthesis: Gulf Research Program

July 10 - July 28, 2017
NCEAS, Santa Barbara, CA

NCEAS
National Center for Ecological Analysis and Synthesis



Open Science for Synthesis: Gulf Research Program

July 10 - July 28, 2017
NCEAS, Santa Barbara, CA

https://nceas.github.io/oss-2017/lessons.html

- Collaboration modes and technologies, virtual collaboration
- Data management, preservation, and sharing
- Data manipulation, integration, and exploration
- Scientific workflows and reproducible research
- Programming using agile and sustainable software practices
- Data analysis and modeling
- Communicating results to broad communities

Today's talk

Our experience using the open science workflow to inform restoration projects in estuaries

Today's talk

Our experience using the open science workflow to inform restoration projects in estuaries

Can we use disparate data to prioritize future restoration projects aimed at improving water quality?

Today's talk

Our experience using the open science workflow to inform restoration projects in estuaries

Can we use disparate data to prioritize future restoration projects aimed at improving water quality?

- ***Synthesize*** data in space and time to evaluate cumulative effects of restoration projects

Today's talk

Our experience using the open science workflow to inform restoration projects in estuaries

Can we use disparate data to prioritize future restoration projects aimed at improving water quality?

- ***Synthesize*** data in space and time to evaluate cumulative effects of restoration projects
- ***Develop*** a decision support tool with empirical observations to evaluate likelihood of potential outcomes

Today's talk

Our experience using the open science workflow to inform restoration projects in estuaries

Can we use disparate data to prioritize future restoration projects aimed at improving water quality?

- ***Synthesize*** data in space and time to evaluate cumulative effects of restoration projects
- ***Develop*** a decision support tool with empirical observations to evaluate likelihood of potential outcomes
- ***Apply*** the tool to guide expectations for future restoration projects

Tampa Bay - from gross to less gross

Past:

- Mid-1970s N load 8.2×10^6 yr^{-1} [Greening and Janicki, 2006]
- Elevated chl-a concentrations
- Increased occurrence of HABs

Tampa Bay - from gross to less gross

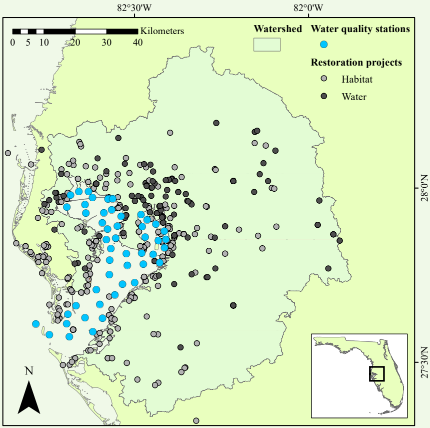
Past:

- Mid-1970s N load 8.2×10^6 yr^{-1} [Greening and Janicki, 2006]
- Elevated chl-a concentrations
- Increased occurrence of HABs

Present:

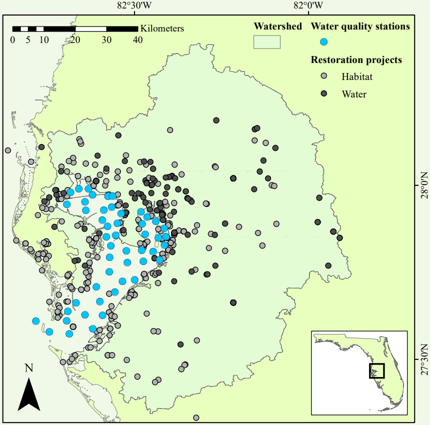
- 2016 seagrass at ˜17k ha [Sherwood et al., 2017]
- Reductions in nutrient load, chlorophyll
- Increase in water clarity [Morrison et al., 2006, Beck et al., 2017]

Tampa Bay - open data sources



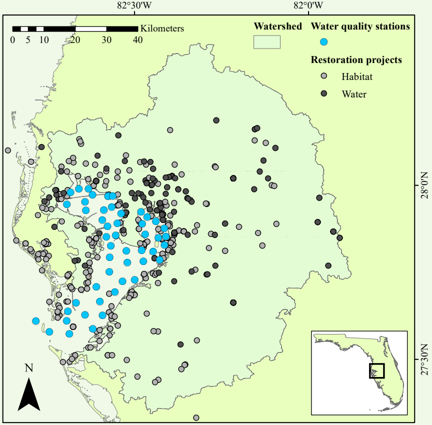
- *Water quality* monitoring dataset: 1974 to present, ~500 obs. per site

Tampa Bay - open data sources



- *Water quality* monitoring dataset: 1974 to present, ~500 obs. per site
- *Restoration projects* dataset: ~500 projects since 1971, habitat and water infrastructure projects

Tampa Bay - open data sources



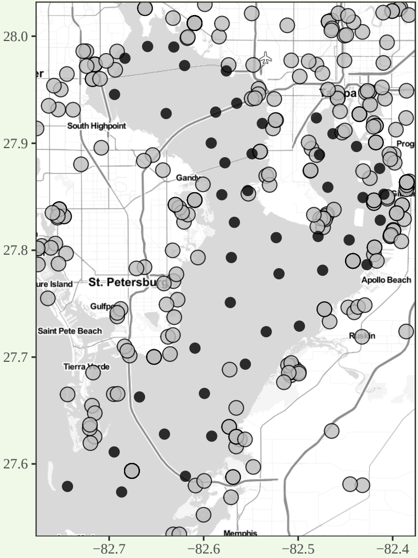
- *Water quality* monitoring dataset: 1974 to present, ~500 obs. per site
- *Restoration projects* dataset: ~500 projects since 1971, habitat and water infrastructure projects

Despite considerable *investments* in restoration, *effectiveness evaluation* continues to elude practitioners at geographic scales [Diefenderfer et al., 2016]

Data munging with open source tools

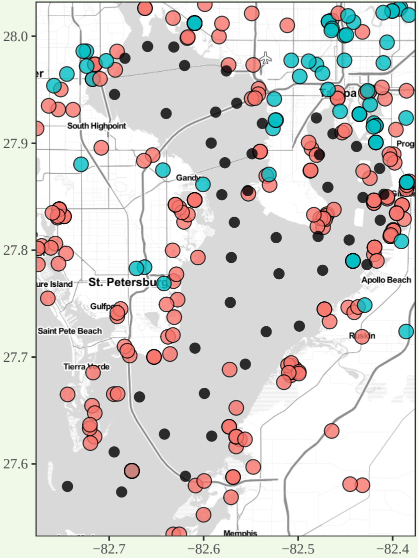
Task 1: Can we empirically link 500 restoration projects to chlorophyll changes at water quality stations over a forty year period?

Data munging with open source tools



Can we link 500 restoration projects to chlorophyll changes over a forty year period?

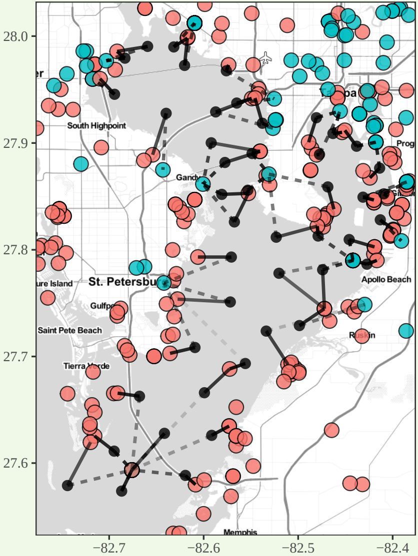
Data munging with open source tools



Can we link 500 restoration projects to chlorophyll changes over a forty year period?

- Consider an effect of restoration ***site type***?

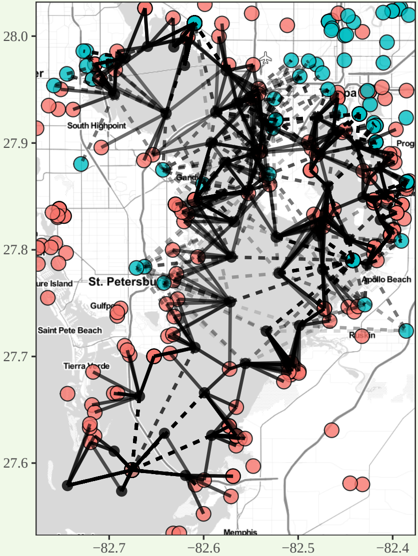
Data munging with open source tools



Can we link 500 restoration projects to chlorophyll changes over a forty year period?

- Consider an effect of restoration ***site type***?
- Consider ***distance*** of sites from water quality stations?

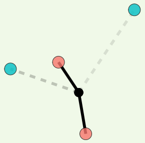
Data munging with open source tools



Can we link 500 restoration projects to chlorophyll changes over a forty year period?

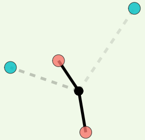
- Consider an effect of restoration ***site type***?
- Consider ***distance*** of sites from water quality stations?
- Consider ***cumulative effects***?

WQ and restoration sites: ***spatial join***

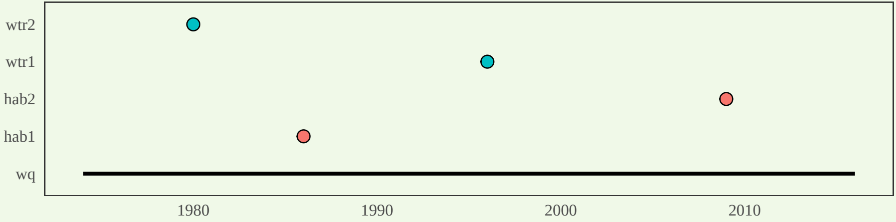


Data munging with open source tools

WQ and restoration sites: *spatial join*

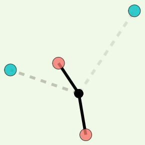


WQ and restoration sites: *temporal join*

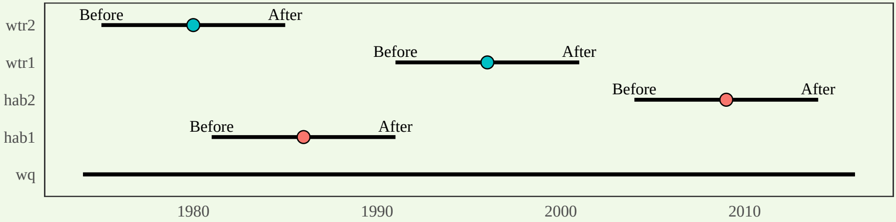


Data munging with open source tools

WQ and restoration sites: *spatial join*

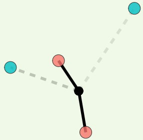


WQ and restoration sites: *temporal join, before/after*

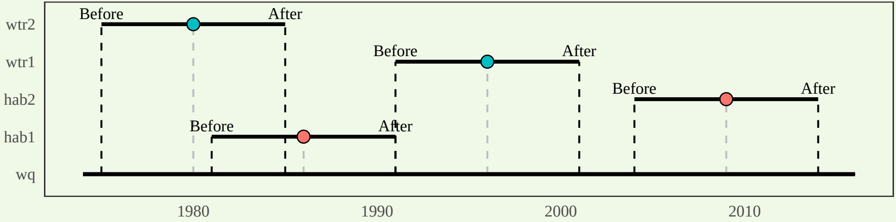


Data munging with open source tools

WQ and restoration sites: *spatial join*

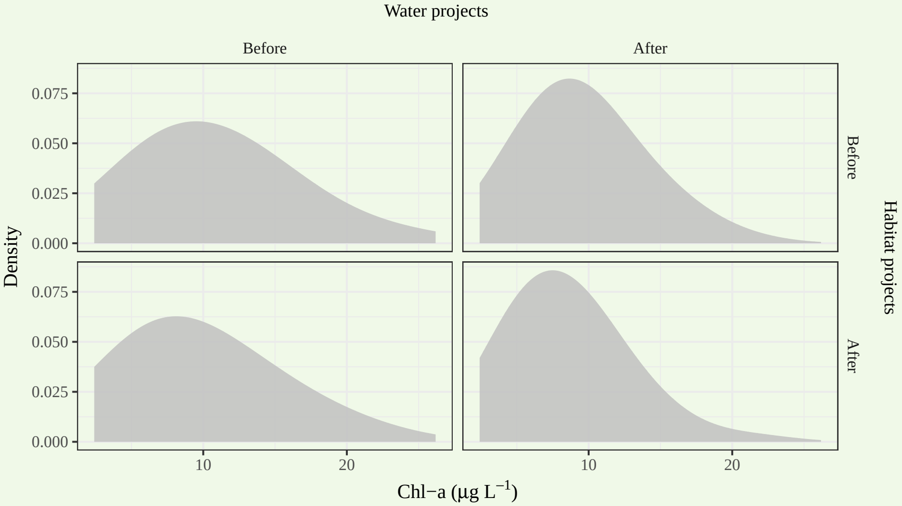


WQ and restoration sites: *temporal join, before/after, slice*



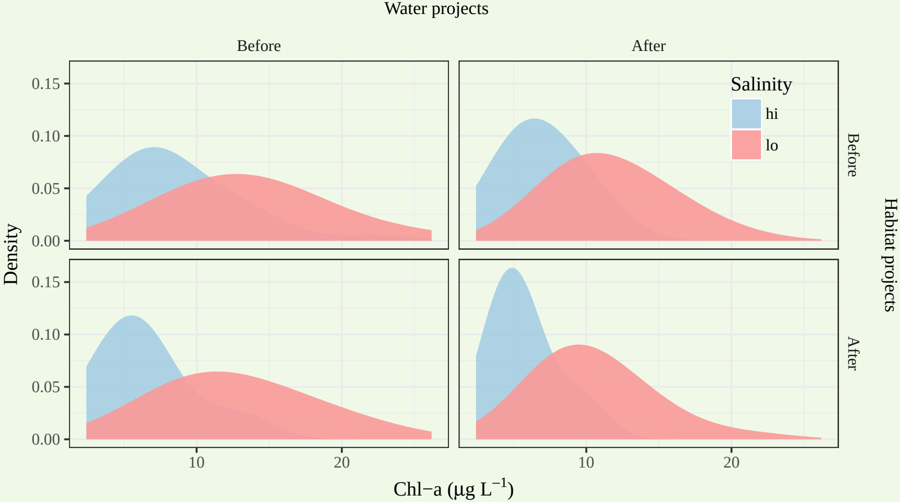
Data munging with open source tools

For *many* water quality stations matched to *many* restoration sites...



Data munging with open source tools

For *many* water quality stations matched to *many* restoration sites...



Data munging with open source tools

What is the probability of low/high chlorophyll given other events?

$$P\left(Chl \mid Event\right) = \frac{P(Event|Chl) \cdot P(Chl)}{P(Event)}$$

Data munging with open source tools

What is the probability of low/high chlorophyll given other events?

$$P\left(Chl \mid Event\right) = \frac{P(Event|Chl) \cdot P(Chl)}{P(Event)}$$

- Do water quality conditions differ by ***restoration type***?
- Does it differ by low/high ***salinity*** as a natural covariate?

Data munging with open source tools

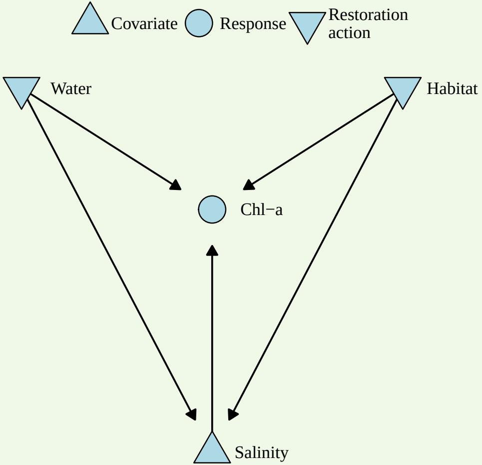
What is the probability of low/high chlorophyll given other events?

$$P\left(Chl \mid Event\right) = \frac{P(Event|Chl) \cdot P(Chl)}{P(Event)}$$

- Do water quality conditions differ by ***restoration type***?
- Does it differ by low/high ***salinity*** as a natural covariate?

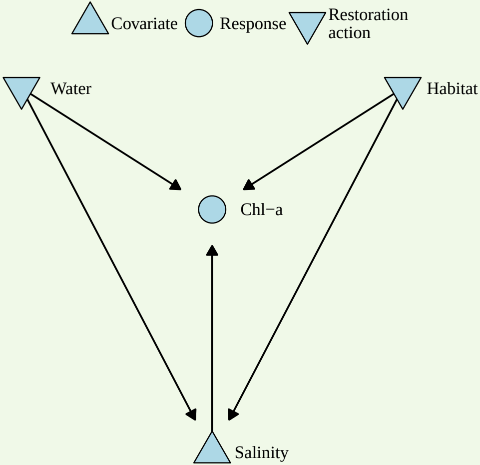
Bayesian models let us evaluate likelihood of potential outcomes given conditional distributions

Using the *conditional probabilities* from the *empirical response*, what's the probability of low chlorophyll...



Data munging with open source tools

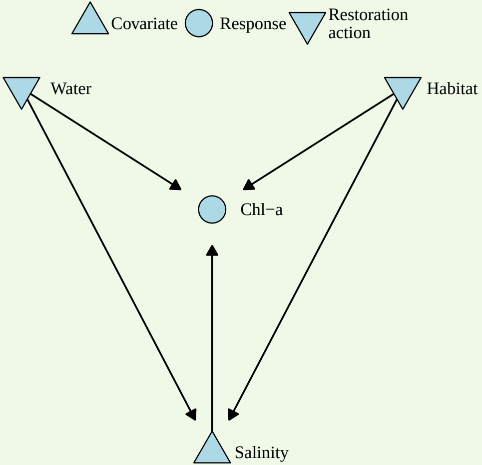
Using the *conditional probabilities* from the *empirical response*, what's the probability of low chlorophyll...



- from water projects

Data munging with open source tools

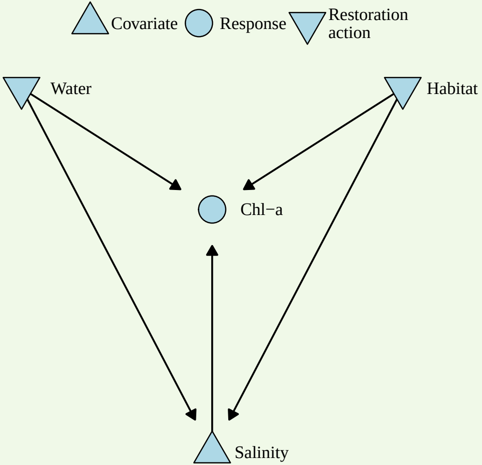
Using the *conditional probabilities* from the *empirical response*, what's the probability of low chlorophyll...



- from water projects
- from habitat projects

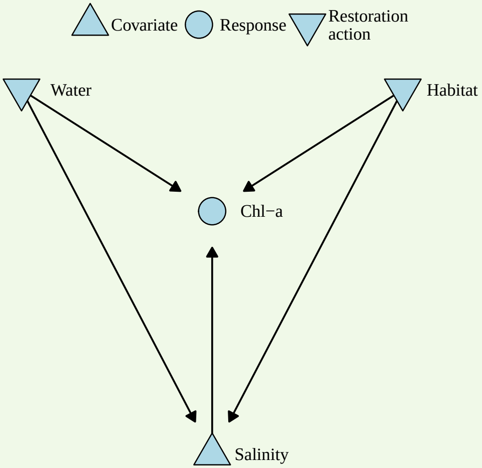
Data munging with open source tools

Using the *conditional probabilities* from the *empirical response*, what's the probability of low chlorophyll...



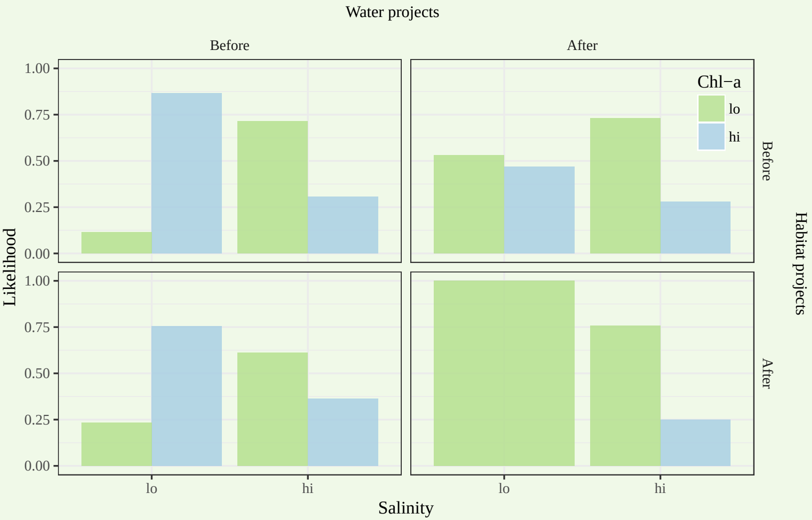
- from water projects
- from habitat projects
- from all projects

Using the ***conditional probabilities*** from the ***empirical response***, what's the probability of low chlorophyll...



- from water projects
- from habitat projects
- from all projects
- by salinity regime

All possible scenarios:



Invidividual scenarios: Probability of **high** chlorophyll

Water projects:

- Before: 53%
- After: 26%

Habitat projects:

- Before: 50%
- After: 50%

Data munging with open source tools

Invidividual scenarios: Probability of **high** chlorophyll

Water projects:

- Before: 53%
- After: 26%

Habitat projects:

- Before: 50%
- After: 50%

Invidividual scenarios: Probability of **low** chlorophyll

Water projects:

- Before: 46%
- After: 77%

Habitat projects:

- Before: 49%
- After: 51%

Data munging with open source tools

Where is the sweet spot?

Probability of low chlorophyll before/after **both project types** by **low/high salinity**

Data munging with open source tools

Where is the sweet spot?

Probability of low chlorophyll before/after **both project types** by **low/high salinity**

Low salinity:

- Before both projects: 13%
- After both projects: 100%

High salinity:

- Before both projects: 69%
- After both projects: 75%

Data munging with open source tools

Take home:

- Water infastructure projects had the largest effect

Data munging with open source tools

Take home:

- Water infastructure projects had the largest effect
- Habitat projects had an effect but less so

Data munging with open source tools

Take home:

- Water infastructure projects had the largest effect
- Habitat projects had an effect but less so
- Effects most noticeable in low salinity zone

Data munging with open source tools

Take home:

- Water infastructure projects had the largest effect
- Habitat projects had an effect but less so
- Effects most noticeable in low salinity zone

But...

Data munging with open source tools

Take home:

- Water infastructure projects had the largest effect
- Habitat projects had an effect but less so
- Effects most noticeable in low salinity zone

But...

- Depends on year, distance combinations

Data munging with open source tools

Take home:

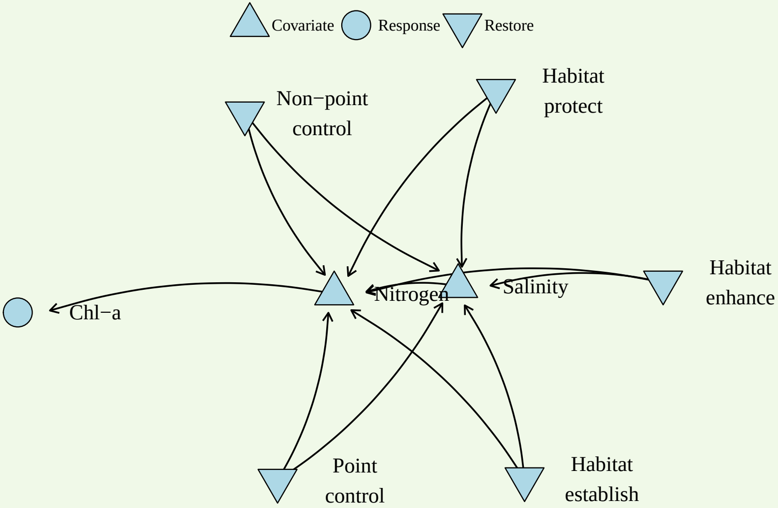
- Water infastructure projects had the largest effect
- Habitat projects had an effect but less so
- Effects most noticeable in low salinity zone

But...

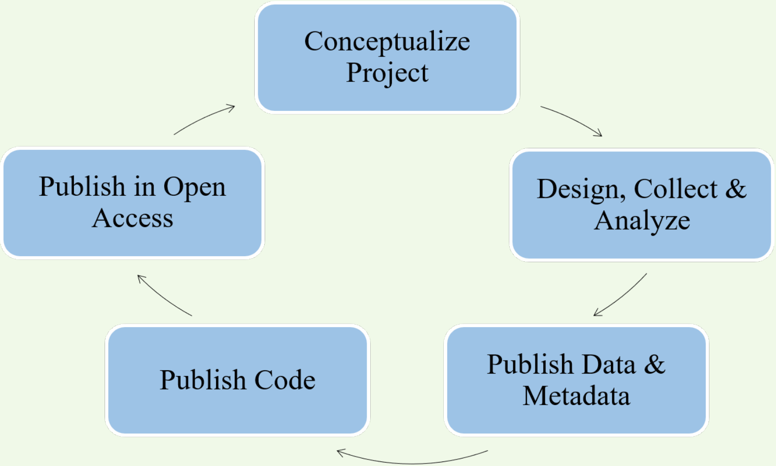
- Depends on year, distance combinations
- This is a simple model

Data munging with open source tools

The holy grail:



Open science workflow



Modified from [Hampton et al., 2015]

Open science workflow

What aspects of our project used and benefitted from open science?

- Early idea conception
- Long distance collaboration
- Transparent and reproducible analysis

Open science workflow

What aspects of our project used and benefitted from open science?

- Early idea conception
- Long distance collaboration
- Transparent and reproducible analysis

... but the circle is not complete

Acknowledgments:

Research staff and employees at NCEAS: M. B. Jones, A. Budden, T. Neal, B. Mecum, C. Lortie, L. Wasser, J. Brun

The Gulf Research Program

Field staff and data managers at Hillsborough County Environmental Protection Commission

Funding sources and contact:

NCEAS
National Center for Ecological Analysis and Synthesis

The National Academies of SCIENCES ENGINEERING MEDICINE
GULF RESEARCH PROGRAM

marcusb@sccwrp.org, 7147553217

GitHub (project): https://github.com/fawda123/restorebayes

GitHub (presentation): https://github.com/fawda123/AWRA_2018

Twitter: @fawda123

References

Beck MW, Hagy III JD, Le C. 2017.
Quantifying seagrass light requirements using an algorithm to spatially resolve depth of colonization.
Estuaries and Coasts, pages 1–17.

Diefenderfer HL, Johnson GE, Thom RM, Buenau KE, Weitkamp LA, Woodley CM, Borde AB, Kropp RK. 2016.
Evidence-based evaluation of the cumulative effects of ecosystem restoration.
Ecosphere, 7(3):e01242.

Greening H, Janicki A. 2006.
Toward reversal of eutrophic conditions in a subtrophical estuary: Water quality and seagrass response to nitrogen loading reductions in Tampa Bay, Florida, USA.
Environmental Management, 38(2):163–178.

Hampton SE, Anderson SS, Bagby SC, Gries C, Han X, Hart EM, Jones MB, Lenhardt WC, MacDonald A, Michener WK, Mudge J, Pourmokhtarian A, Schildhauer MP, Woo KH, Zimmerman N. 2015.
The tao of open science for ecology.
Ecosphere, 6(7):1–13.

Morrison G, Sherwood ET, Boler R, Barron J. 2006.
Variations in water clarity and chlorophyll a in Tampa Bay, Florida, in response to annual rainfall, 1985-2004.
Estuaries and Coasts, 29(6):926–931.

Sherwood ET, Greening HS, Johansson JOR, Kaufman K, Raulerson G. 2017.
Tampa Bay (Florida, USA): Documenting seagrass recovery since the 1980s and reviewing the benefits.
Southeastern Geographer, 57(3):294–319.