

Oil impacts and open science

Introducing the Open Derby

Ensuring openness and reproducibility in ecology is an ongoing challenge, although open tools (R Studio, Github) are free to use, and user support is everywhere. To illustrate how ecology collaborations can fit within an open science framework, we had the idea of running an ‘Open Derby’. Similar to a research derby, where grad students and post-docs from different academic departments and universities collaborate together to solve a conservation problem, our idea extends the derby concept to working fully openly: learning version control in GitHub, writing and analysing data in RMarkdown, and publishing all project development in an open notebook. Our end goal is to examine a conservation issue using open access (OA) data, and publish our findings and code in an OA journal. Ultimately, we hope to share Open Derby with other graduate students and academics across the life sciences, with the hope that we can provide a model for switching entire lab groups into open and reproducible research machines.

If you are not familiar with open science tools, RMarkdown is a wonderful method of writing code into a manuscript that can be formatted into manuscripts, reports, and presentations. We use RMarkdown in tandem with GitHub, a service that allows version control of your data, analyses, and manuscripts, and is incredibly useful for managing input between collaborators. Learning these tools can be daunting but learning with friends makes it easier and more fun to build these skills. We are six PhD students from 4 institutions based in 3 countries, who are present or past members of Dr. Julia Baum’s marine ecology lab at the University of Victoria and have come together to demo the Open Derby concept:

James Robinson: PhD student in the Baum lab, studying macroecology of tropical Pacific coral reefs (defending April 2017) and leader of the Open Derby project

Danielle Claar: PhD student in the Baum Lab, studying the dynamics of coral symbioses in response to environmental stressors

Easton White: PhD student at the University of California, Davis. He uses mathematical and statistical techniques to address various ecological and conservation questions.

Jillian Dunic: PhD student at Simon Fraser University, studying the effect of multiple stressors on tipping points in eelgrass.

Jamie McDevitt-Irwin: MSc student in the Baum lab evaluating how stressors, specifically human disturbance and bleaching stress from an El Niño event, influence the community structure of the coral microbiome (defending April 2017).

Geoffrey Osgood: PhD student in the Baum lab, using data and models to study shark populations in South Africa and on Pacific reefs

In January 2017, we met (virtually) to decide on a research question for Open Derby. Our aim was to address a question that would be accessible under an open science platform (e.g. open source data). And **SPOILER**, we failed in our first attempt. In the interests of openness, and publishing null results, here is what we did:

Our conservation question - quantifying oil spill impacts across regions

In light of recent political moves towards greater oil extraction in N America, we decided to examine oil spill impacts across the west coast of North America. We focused on shellfish fisheries, reasoning that sessile organisms would experience high mortality soon after an oil spill (e.g. shellfish, often mussels and more mussels), and that subsequent fisheries closures (e.g. English Bay, BC, Canada, Gulf of Mexico, USA, Santa Barbara, CA, USA) would instantly impact catch.

Our workflow

Tools used: Rmarkdown, Github, Slack, etherpad, Skype

As mentioned earlier, we used Rmarkdown and Github to manage our workflow. We met once every one or two weeks on Skype to discuss project goals and delegated tasks, taking notes on Skype calls in a publicly accessible etherpad. These meetings, combined with Github, allowed us to divide and conquer without duplicating each other's work. The goal was to have everyone only spend 1-3 hours on the project per week (after all, we each have dissertations to write). Everyone contributed to coding in R—whether it was data cleaning, analysis, or visualizations—and writing the initial manuscript. We also used Slack to regularly chat about the project in between Skype meetings. Slack eliminates the need for endless email chains and also integrates with Github, allowing participants to be notified when a repository change has occurred. In the future, we might take notes on a markdown file that can be easily version controlled, and may make use of Github's wiki and project planning features.

Open datasets - oil spills and fisheries catch

We extracted oil spill data from the Emergency Response Division of the National Oceanic and Atmospheric Administration (NOAA; <https://incidentnews.noaa.gov/raw/index>) which recorded all major spills that occurred between 1967-2016 in marine environments on the west coasts of U.S.A. and Canada (Figure 1). However, we were unable to find oil spill data from British Columbia (BC). After extensive web searches, emails and even a freedom of information request, we were told there is no existing electronic record of BC coastal oil spills.

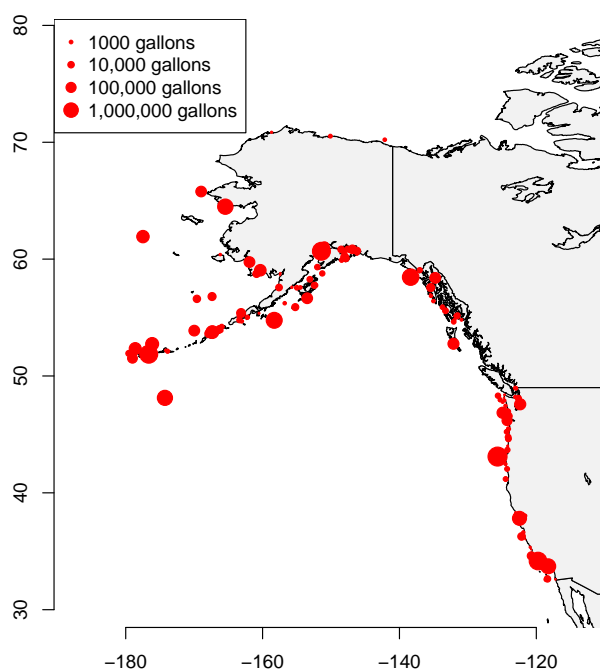


Figure 1 Oil spill location and size across west coast of North America from 1967-2016 included in NOAA's incident news dataset.

We extracted fisheries catch data from the Sea Around Us Project (SaU), which collates global, spatially-explicit estimates of annual fisheries catch at the national level (Pauly & Zeller 2015; Zeller et al. 2016). The SaU dataset uses international and national catch statistics to generate taxon-level catch estimates at a 0.5 degree resolution, spanning 1950-2013, and provided us with California, Oregon, Washington, BC, and Alaska fisheries catch. Yet, as promising as this data appeared at first, we were advised not use this data

at spatial scales below the exclusive economic zone (EEZ) level, as data are aggregated. We also contacted Canadian and US fisheries departments to request port-level catch data.

Sadly, data resolution issues sunk our project. With fishing catch data, smaller spatial aggregations of the SaU dataset were simply EEZ level trends with catch estimates scaled downwards (Figure 2), and we did not reasonably expect to detect oil spill impacts at the EEZ level. We were also unable to locate consistent port-level data that spanned either BC, California, or Alaska. Similarly, the oil spill incidents were not consistently reported, and it was unclear how spill estimates were generated - a contact from NOAA warned us that there was no established methodology for estimating the size of an oil spill, and many reported incident volumes involve some amount of guesswork.

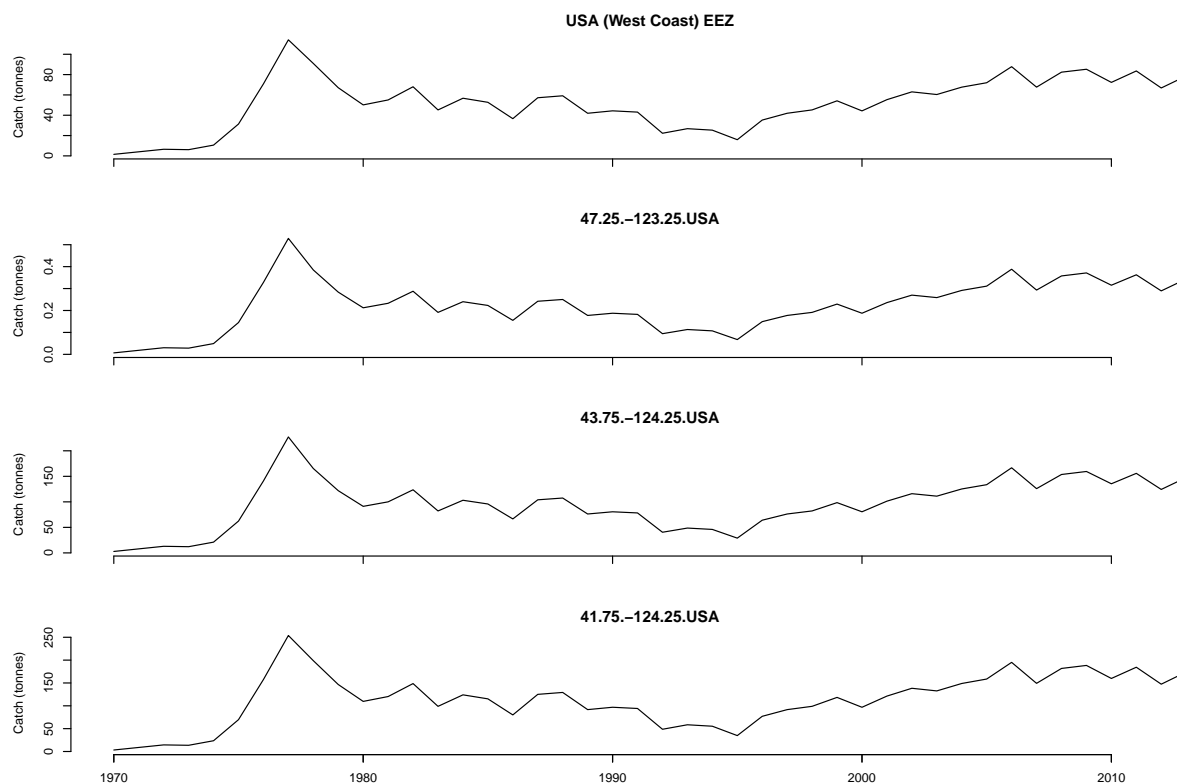


Figure 2 Catch estimates for USA west coast EEZ (top panel) and 3 randomly selected 0.5 degree resolution grid cells (bottom 3 panels, labelled with lat-lon centroid of grid cell) from 1970-2010.

Beyond data issues, the ecological impact of an oil spill is highly context-dependent, and will be affected by the environmental conditions, clean-up response, and type of oil contaminant. As a result, our understanding of oil spill impacts on marine ecosystems is, sadly, limited to large, catastrophic events. We urge more openness and data collection of spill characteristics, so that researchers may begin to quantify the cumulative impacts of smaller, more frequent spills on marine ecosystems.

Future Work

You can find our ideas, data, and progress in our public Github repository. We welcome any suggestions that might help us to resurrect this project, and hope to inspire others to write up their failed projects. Our project lead - James Robinson - is currently developing Open Derby for general consumption as a member of Mozilla's Open Science Leadership training programme.

But, we are now brainstorming conservation questions for our next attempt. Ecologists, please, send us your unanswered conservation problems, unanalyzed datasets, and suggestions for advancing open ecology! This time we will each bring a dataset and question to the brainstorming session. We will let you know how we progress!

James Robinson, Danielle Claar, Jillian Dunic, Jamie McDevitt-Irwin, Geoffrey Osgood, Easton White