Improving the communication and transparency of stock assessment using interactive visualization tools

Paul M Regular, Gregory J Robertson, etc.?

Fisheries and Oceans Canada, Northwest Atlantic Fisheries Center, 80 East White Hills, St. John’s, Newfoundland and Labrador, A1C 5X1, Canada

# Introduction

In the age of data-intensive science and big data, scientists across many fields are faced with the challenge of synthesizing and communicating information from large and complex data sets. Scientific data are often presented as tabular summaries and static plots, however, seeing patterns in tables can be difficult and static plots preclude detailed explorations of the underlying data. New open source tools are bridging this gap by making it easier to build interactive figures (Perkel 2018). In particular, browser-based data visualizations have been playing an increasing prominent role in communicating information on a wide range of topics. Interactive graphics are frequently utilized by media outlets, such as the FiveThirtyEight, and scientists from a wide range of disciplines are starting to apply interactive visualization tools to explore and communicate their results (Yeatman et al. 2018; Tushar and Reich 2017; Wick et al. 2015). This trend has been fueled by the development of JavaScript libraries (e.g. D3 (Bostock, Ogievetsky, and Heer 2011), plotly (Inc. 2015), leaflet (Leaflet 2015), highcharts (Highsoft 2015)) that enable the production of dynamic and interactive data visualizations in modern web-browsers. Interfaces to JavaScript libraries have also been developed to allow interactive plots to be coded using commonly used languages such as R (R Core Team 2017) or Python (Foundation 2010). By laying the technical infrastructure for creating interactive graphics, these new tools are ushering in a new era of Big Data by facilitating the exploration and communication of complex data and models.

The field of fisheries research has a long history of collecting and synthesizing a plethora of data using state-of-the-art computational technologies (Megrey and Moksness 2008). Fisheries science has evolved alongside the rise of the computing era, and as tools improve so does the management advice provided by fisheries scientists. This management advice generally culminates from a stock assessment process whereby various sources of data and statistical methods are used to model the history, current status and future status of one or more fish stocks (Hilborn and Walters 1992). Recent advances in computational technologies have streamlined the process of stock assessment by facilitating the development and application of integrated assessment models (Maunder and Punt 2013). This means that analyses that were once carried out independently on different data sets can now be integrated into one analysis by using statistical modeling tools such as JAGS (Plummer and others 2003), AD Model Builder (Fournier et al. 2012) and Template Model Builder (Kristensen et al. 2015). From a computational perspective, analyses of a variety of large data sets has never been easier. However, from a human perspective, contemporary stock assessment biologists are faced with the challenge of managing data from a variety of sources and also understanding the algorithms that convert these data to advice for fisheries managers. Here we aim to demonstrate how interactive visualization tools provide an efficient and effective means of exploring and communicating the ever expanding array of data inputs and model outputs. First, we focus on data that are commonly used in stock assessment and provide one example where interactive maps were used to simplify the detailed exploration of data from a long-term tagging program. Second, we focus on the modeling aspect of stock assessment and, again, we use an example to demonstrate how dynamic and interactive data visualizations can be used to explore, diagnose and communicate results from an integrated assessment model. This structure corresponds to the two general steps in the stock assessment process: 1) knowing the data, and 2) data modeling.

## Knowing the data

The data sets used in stock assessments are constantly growing. This growth in data either stems from the continuation of long-term monitoring efforts or from the addition of new monitoring programs. As such, stock assessment biologists often have to manage large volumes of data from a variety of sources. For instance, time series of reported landings and catch-at-age are “fishery-dependent” data that are frequently used in stock assessments. These data are often analyzed in conjunction with data from “fishery-independent” surveys that track changes in abundance and, in many cases, also monitor trends in biological factors such as age composition, growth rates, sex ratios and maturation stages. For some data-rich stocks, mark and recapture studies are also carried out to estimate movement, migration, growth rate, natural mortality, and discard mortality. All of the above-mentioned data sets are multidimensional and as the volume and variety of these data increases, it becomes more difficult to be aware of the details of each data set and discover key patterns within each. This challenge can be mediated, to a degree, by the application of interactive visualization tools as that they allow the construction of visual representations of the data that demonstrates large-scale patterns while also allowing the detailed exploration of the underlying data. For instance, the ability to zoom in on features or areas of interest and hover over specific points to reveal more information creates an interactive user-driven experience that expedites explorations of the data. This is exemplified by a recently developed interactive mapping tool developed for the exploration of a long-term tagging study.

* *Greg, can you introduce this tool, providing some info on the tagging database and the tools used to make the tool (Shiny, leaflet)?*
* *TODO: the intro paragraph to this section could use some references if it is necessary*

## Data modeling

Synthesizing data from multiple sources presents a key challenge to stock assessment. Analyses of different data sources were traditionally carried out independently and the summaries or parameters from these analyses were used in the assessment model. This approach, however, is less than ideal because information may be lost and uncertainty may be unaccounted for when we “do statistics on the statistics” (Link 1999; Maunder and Punt 2013). Such issues have largely been curtailed in contemporary stock assessments thanks to advances in software that have facilitated the analysis of all available data, in as raw a form as appropriate, in a single integrated analysis (Maunder and Punt 2013). Specifically, the application of statistical modeling tools such as JAGS (Plummer and others 2003), AD Model Builder (Fournier et al. 2012) and Template Model Builder (Kristensen et al. 2015) allow the construction of a joint likelihood for an array of observations to, in theory, extract as much information as possible about the biological and fishery processes. However, integrated analyses are not a panacea because model misspecifications and data conflicts are an inevitable consequence of simplifying reality to a small series of equations (Maunder and Piner 2017). A potential solution to this quandary to use a superensemble model, whereby multiple models with different structures are run and their predictions are supplied as covariates to an additional superensemble model (Anderson et al. 2017). Ensemble approaches reduce the risk of picking the “wrong” model and also expands the range of hypotheses explored (Dietterich 2000). These advances greatly improve our ability to assess the status and trends of fish populations, however, modern stock assessment biologists are now faced with the overwhelming task of understanding an ever expanding array of data inputs and model outputs.

* *TODO: tweak this intro paragraph and work up sugue to NCAM and the dashboard*

# Other draft text

A wide array data may be collected for an assessment and, formally, the process often reduces to algorithms that convert these data to advice for fisheries managers. In some cases, particularly for commercially valuable species, this means that hundreds or thousands of historical data points from the monitoring program of a stock gets reduced into a single policy value, such as a recommended catch quota (Maunder, Schnute, and Ianelli 2009). Sound management advice therefore hinges on the ability of stock assessment biologists to manage, interpret and communicate large volumes of data from a variety of sources.

In this data-intensive era of stock assessment, one of the difficulties becomes visualizing the inputs and outputs to integrated stock assessment models. The standard approach has been to create a series of tables and static plots to help assess the inputs and model fits. Data presented in tables is incredibly valuable, but as human beings, seeing patterns can be challenging. Static figures, in contrast, help reveal patterns but they are divorced from the underlying data and, as such, limit detailed explorations. New tools are bridging this gap by making it easier to build interactive figures (Perkel 2018).

Interactive visuals…visual representation of the data and model that help us interpret complex data at a glance <https://campus.sagepub.com/blog/3-benefits-of-interactive-visualization>

<https://centricdigital.com/blog/digital-trends/the-value-of-interactive-data-visualization-and-why-it-matters-to-business/>

<https://www.nature.com/articles/s41467-018-03297-7>

<https://esajournals.onlinelibrary.wiley.com/doi/full/10.1890/120103>

<https://www.nature.com/articles/d41586-018-01322-9>

“The ultimate objective of Fisheries Science is to inform management” (<http://www.fao.org/tempref/docrep/fao/008/a0212e/a0212E02.pdf>)

# References

Anderson, Sean C, Andrew B Cooper, Olaf P Jensen, Cóilín Minto, James T Thorson, Jessica C Walsh, Jamie Afflerbach, et al. 2017. “Improving Estimates of Population Status and Trend with Superensemble Models.” *Fish and Fisheries* 18 (4). Wiley Online Library: 732–41.

Bostock, Michael, Vadim Ogievetsky, and Jeffrey Heer. 2011. “D Data-Driven Documents.” *IEEE Transactions on Visualization & Computer Graphics*, no. 12. IEEE: 2301–9.

Dietterich, Thomas G. 2000. “Ensemble Methods in Machine Learning.” In *International Workshop on Multiple Classifier Systems*, 1–15. Springer.

Foundation, Python Software. 2010. “Python Language Reference, Version 2.7.” Python Software Foundation Wilmington, DE.

Fournier, David A, Hans J Skaug, Johnoel Ancheta, James Ianelli, Arni Magnusson, Mark N Maunder, Anders Nielsen, and John Sibert. 2012. “AD Model Builder: Using Automatic Differentiation for Statistical Inference of Highly Parameterized Complex Nonlinear Models.” *Optimization Methods and Software* 27 (2). Taylor & Francis: 233–49.

Highsoft. 2015. <https://www.highcharts.com/>.

Hilborn, Ray, and Carl J Walters. 1992. *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. Chapman; Hall.

Inc., Plotly Technologies. 2015. “Collaborative Data Science.” Montreal, QC: Plotly Technologies Inc. <https://plot.ly>.

Kristensen, Kasper, Anders Nielsen, Casper W Berg, Hans Skaug, and Brad Bell. 2015. “TMB: Automatic Differentiation and Laplace Approximation.” *arXiv Preprint arXiv:1509.00660*.

Leaflet. 2015. <https://leafletjs.com/>.

Link, William A. 1999. “Modeling Pattern in Collections of Parameters.” *The Journal of Wildlife Management*. JSTOR, 1017–27.

Maunder, Mark N, and Kevin R Piner. 2017. “Dealing with Data Conflicts in Statistical Inference of Population Assessment Models That Integrate Information from Multiple Diverse Data Sets.” *Fisheries Research* 192. Elsevier: 16–27.

Maunder, Mark N, and André E Punt. 2013. “A Review of Integrated Analysis in Fisheries Stock Assessment.” *Fisheries Research* 142. Elsevier: 61–74.

Maunder, Mark N, Jon T Schnute, and James N Ianelli. 2009. “Computers in Fisheries Population Dynamics.” In *Computers in Fisheries Research*, 337–72. Springer.

Megrey, Bernard A, and Erlend Moksness. 2008. *Computers in Fisheries Research*. Springer.

Perkel, Jeffrey M. 2018. “Data Visualization Tools Drive Interactivity and Reproducibility in Online Publishing.” *Nature* 554 (7690): 133–34.

Plummer, Martyn, and others. 2003. “JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling.” In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Vol. 124. 125.10. Vienna, Austria.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Tushar, Abhinav, and Nicholas G Reich. 2017. “Flusight: Interactive Visualizations for Infectious Disease Forecasts.” *Journal of Open Source Software* 2 (13). NIH Public Access.

Wick, Ryan R, Mark B Schultz, Justin Zobel, and Kathryn E Holt. 2015. “Bandage: Interactive Visualization of de Novo Genome Assemblies.” *Bioinformatics* 31 (20). Oxford University Press: 3350–2.

Yeatman, Jason D, Adam Richie-Halford, Josh K Smith, Anisha Keshavan, and Ariel Rokem. 2018. “A Browser-Based Tool for Visualization and Analysis of Diffusion Mri Data.” *Nature Communications* 9 (1). Nature Publishing Group: 940.