The Journal of Open Source Software

dia: An R package for the National Oceanic and Atmospheric Administration dam impact analysis

Daniel S. Stich[1¶], **Julie L. Nieland**[2], and **Timothy F. Sheehan**[2]

1 Biology Department and Biological Field Station, State University of New York at Oneonta, NY 13280 USA **2** National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northeast Fisheries Science Center, Woods Hole, MA 02543 USA ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

Populations of anadromous (sea-run) fishes such as Atlantic salmon *Salmo salar* have experienced severe global declines as a result of pollution, overfishing, and construction of dams (Limburg & Waldman, 2009). Life-history-based simulation models are commonly used for planning and implementing fisheries recovery activities for many diadromous species (Barber et al., 2018; e.g., Nieland et al., 2013; Stich et al., 2019; Zydlewski et al., 2021). Before open-source tools became prevalent, many institutionalized decision-support tools utilizing such models relied on closed-source or paid software. For example, the National Oceanic and Atmospheric Administration (NOAA) Dam Impact Analysis (DIA) was originally created as a stochastic life cycle model for Atlantic salmon in the `@RISK` add-in within Microsoft Excel (Nieland et al., 2013). This tool differs from those hosted in similar R packages for non-salmonid species (e.g. Stich et al., 2019; Zydlewski et al., 2021) in that it incorporates homing behavior (probability of adults returning to natal streams or straying to others) and integrates results of physical modeling to inform population dynamics (Nieland et al., 2013; Nieland & Sheehan, 2020). This class of tools, in general, provides advantages for decision making related to anadromous species because it allows integration of geographically and temporally explicit stock dynamics (e.g., influences of dams) that are not readily implemented in classical fisheries stock assessment tools such as those available in existing R packages (Erickson et al., 2022; e.g. Kell et al., 2007; Ogle et al., 2022). We created the `dia` package (Stich et al., 2021) for the R programming language (R Core Team, 2024) as a freely accessible, open-source implementation of these tools that will promote transparency in planning and decision making.

Statement of need

`dia` is an R-based implementation of a previously closed-source life cycle model of Atlantic salmon population dynamics that is used to understand population response to natural and anthropogenic influences in freshwater and marine environments. It was developed to assess the sensitivity of restoration outcomes to uncertainty in life-history inputs alongside the impacts of dams and restoration activities. The DIA model uses empirical life-history estimates (e.g., survival), predictive flow and resulting flow-specific dam survival modeling, and other empirical data in freshwater and marine environments to simulate consecutive generations of Atlantic salmon in the Penobscot River (Maine, USA) under varying environmental conditions or management decisions. As the largest remaining population of this critically endangered species in the USA, the population is intensively managed. Management decisions include fish passage rates at dams, fishery harvest rates, and numbers and locations for hatchery stocking of fish (Nieland et al., 2013). Since development, it has been used for mechanistic exploration of key life-history uncertainties within the context of species recovery (Nieland et al., 2015) and to support decision-making at federally regulated hydropower dams on the Penobscot River (e.g., National Marine Fisheries Service, 2013; Nieland & Sheehan, 2020).

We created `dia` for use by fisheries researchers, managers, and practitioners interested in understanding population dynamics of intensively managed endangered Atlantic salmon in the USA. The R package maintains the core routines from the original closed-source version of the model by replicating spreadsheet-based calculations, and incorporates original data and parameter sets as built-in objects that serve as default values for arguments of the primary user-facing functions. However, it also allows exploration of uncertainty associated with life-history parameters and investigation of future restoration scenarios through a variety of user-facing options.

The two primary user-facing functions within the `dia` package are `run_dia()` and `run_dia_shiny()`, which provide redundant interfaces for using Dam Impact Analysis (DIA) models in different ways. The `run_dia()` function provides an extensible interface to DIA. It can be used for long-run simulation or decision-optimization studies. It allows incorporation of user-specified data sets such as flow-correlated survival probabilities at dams and in free-flowing river reaches, marine survival and other life-history inputs, or fish-stocking data. The `run_dia_shiny()` function deploys a graphical user interface using the `shiny` package (Chang et al., 2024) that is less extensible but more easily used by fishery managers and practitioners who may be less familiar with programming and it also includes exportable results from simulation models including `.csv` or other flat-file formats and default plots through the `ggplot2` R package [Wickham (2016); Wickham et al. (2019); Figure 1]. Both can be deployed on networked servers as other R or `shiny` applications to improve accessibility or facilitate use on high performance computers for large simulations. The GitHub repository (Stich et al., 2021) includes additional instructions for installation and a variety of potential uses of `run_dia()` and `run_dia_shiny()`, with shorter examples included in the package help files. While implementation is currently limited to the Penobscot River as a priority conservation water in the USA, the package serves as one example to help generalize these modeling approaches to Atlantic salmon and other sea-run fish in watersheds globally. Specifically, while many of the built-in datasets and helper functions in `dia` are generalized or could be used to simulate life-history information for other systems and species, generalizing the geographic component (i.e., structural river system) represents an important priority for future development.

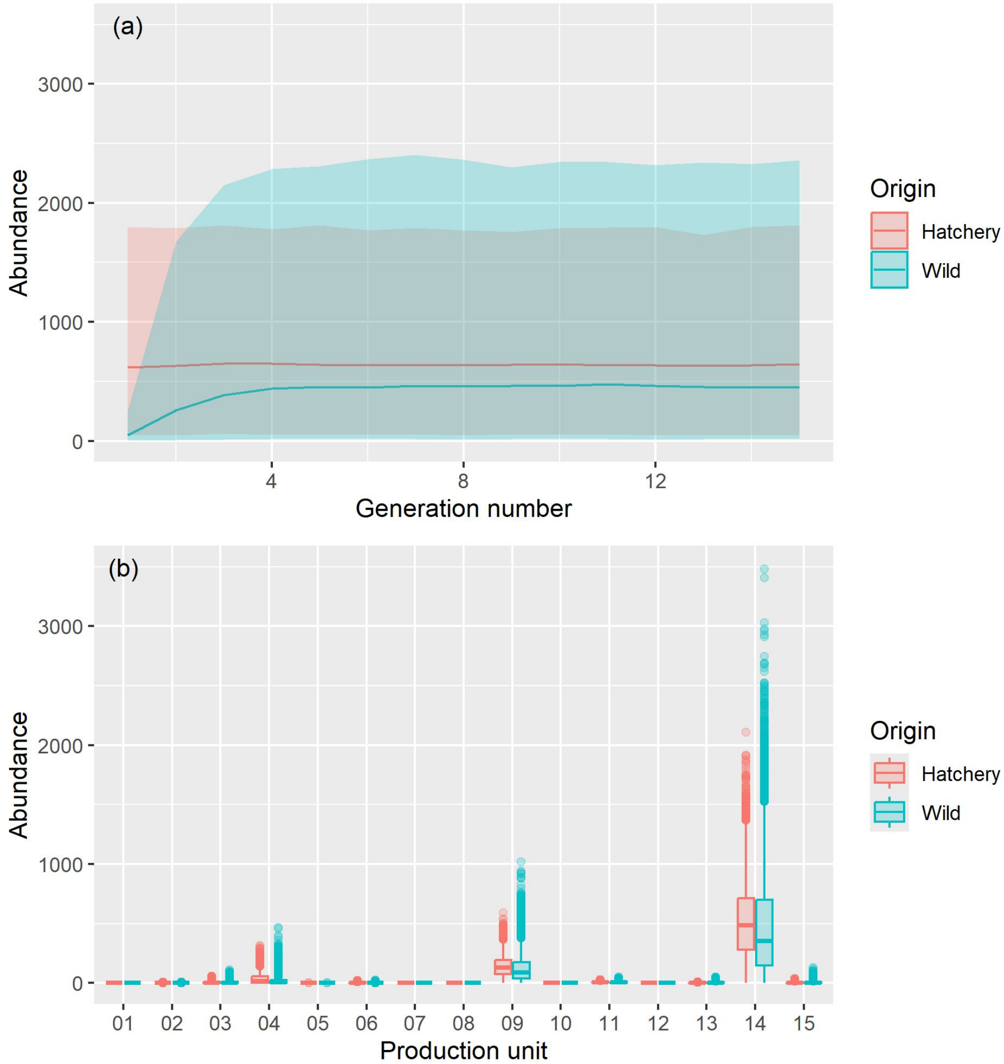

Figure 1: Example graphical outputs using the default argument values (Nieland & Sheehan, 2020) for `run_dia()` to run 10,000 simulations, showing (a) the number of two-sea-winter adult females of hatchery or wild origin returning to the watershed during each generation to spawn, and (b) the number of those fish returning to each production unit within the watershed after 15 generations.

Acknowledgements

We acknowledge contributions from Rory Saunders, Tara Trinko-Lake, Jeffrey Murphy, and Justin Stevens in the original development of the NOAA Dam Impact Analysis model. This work was funded by the NOAA National Marine Fisheries Service Northeast Fisheries Science Center (award number NMFS-NEFSC-199 MOU), with additional support from the SUNY Oneonta Biological Field Station. We thank the reviewers for helpful guidance and comments that greatly improved the quality of this paper and the open-source software package. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Barber, B. L., Gibson, A. J., O'Malley, A. J., & Zydlewski, J. (2018). Does what goes up also come down? Using a recruitment model to balance alewife nutrient import and export. *Marine and Coastal Fisheries*, *10*(2), 236–254. https://doi.org/10.1002/mcf2.10021

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2024). *Shiny: Web application framework for r*. https://doi.org/10.32614/CRAN.package.shiny

Erickson, R. A., Stich, D. S., & Hebert, J. L. (2022). fishStan: Hierarchical bayesian models for fisheries. *Journal of Open Source Software*, *7*(71), 3444. https://doi.org/10.21105/joss.03444

Kell, L. T., Mosqueira, I., Grosjean, P., Fromentin, J.-M., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M. A., Poos, J. J., Scott, F., & Scott, R. D. (2007). FLR: An open-source framework for the evaluation and development of management strategies. *ICES Journal of Marine Science*, *64*, 640–646. https://doi.org/10.1093/icesjms/fsm012

Limburg, K. E., & Waldman, J. R. (2009). Dramatic declines in North Atlantic diadromous fishes. *BioScience*, *59*(11), 955–965. https://doi.org/10.1525/bio.2009.59.11.7

National Marine Fisheries Service. (2013). *National Marine Fisheries Service Endangered Species Act Biological Opinion Amendment of License for the Mattaceunk Project f/NER/2013/9640*. National Oceanic and Atmospheric Administration. https://repository.library.noaa.gov/view/noaa/55737

Nieland, J. L., & Sheehan, T. F. (2020). Quantifying the effects of dams on Atlantic salmon in the Penobscot River watershed, with a focus on Weldon Dam. In *Northeast Fisheries Science Center Reference Document 19-16*. National Oceanic and Atmospheric Administration. https://doi.org/10.25923/v67x-kk62

Nieland, J. L., Sheehan, T. F., & Saunders, R. (2015). Assessing demographic effects of dams on diadromous fish: a case study for Atlantic salmon in the Penobscot River, Maine. *International Council for the Exploration of the Sea Journal of Marine Science*, *72*(8), 2423–2437. https://doi.org/10.1093/icesjms/fsv083

Nieland, J. L., Sheehan, T. F., Saunders, R., Murphy, J. S., Trinko Lake, T., & Stevens, J. R. (2013). Dam impact analysis model for Atlantic salmon in the Penobscot River, Maine. In *Northeast Fisheries Science Center Reference Document 13-09*. National Oceanic; Atmospheric Administration. https://repository.library.noaa.gov/view/noaa/4559

Ogle, D. H., Doll, J. C., Wheeler, P., & Dinno, A. (2022). *FSA: Fisheries stock analysis*. https://doi.org/10.5281/zenodo.597719

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Stich, D. S., Nieland, J. L., & Sheehan, T. F. (2021). dia: Atlantic salmon dam impact analysis (DIA) for R. In *GitHub repository*. GitHub. https://doi.org/10.5281/zenodo.13376045

Stich, D. S., Sheehan, T. F., & Zydlewski, J. D. (2019). A dam passage performance standard model for American shad. *Canadian Journal of Fisheries and Aquatic Sciences*, *76*(5), 762–779. https://doi.org/10.1139/cjfas-2018-0008

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://doi.org/10.32614/CRAN.package.ggplot2

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemun, G., Haye, A., Henr, L., Heste, J., Kuh, M., Lin Pederse, T., Mille, E., Bach, S. M., Müll, K., Jeroen Oo, D., Seid, D. P., V., … Yutani, H. (2019). Welcome to the tidyverse.

Journal of Open Source Software, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Zydlewski, J., Stich, D. S., Roy, S., Bailey, M., Sheehan, T., & Sprankle, K. (2021). What have we lost? Modeling dam impacts on American shad populations through their native range. *Frontiers in Marine Science*, *8*. https://doi.org/10.3389/fmars.2021.734213