RSPARROW v2.1 review

Marcus W. Beck

Mar 21, 2025

The following is a code and domain review of *RSPARROW* prior to release of version 2.1. The review is based off all content on the branch `rmdremove` of the `rsparrow-development` repository. The branch was downloaded on March 20th, 2025 and unzipped to a local directory, after which an RStudio Project (v2024.12.1 build 563) was created in the same directory for evaluating the model. The package was loaded using `devtools::load_all('RSPARROW_master')` and any missing package dependencies were installed following the prompt, all of which installed without issue. A local installation of R version 4.4.2 was used, which was also the recommended version included with the repository.

As with my previous review of version 2.0, this review was created after running the calibrated total phosphorus model for Tampa Bay using the file *UserTutorialDynamic/results/sparrow_control.R*. The model ran successfully without issue and results were available after a few minutes, both in the `results` folder and through the Shiny DSS.

Code review

My previous review noted that several .Rmd files were located in the R folder of the package. As noted, only R files should be at this location and all of the .Rmd files have been removed from the package. Reviewing the content in the posted issue in the repository outlines these changes and I understand the functionality provided by the .Rmd files has been converted to separate R functions. I know this is a tedious way to create .Rmd reports, but I'm glad this change has been made given that debugging and additional modifications will be much easier. I greatly appreciate that this change has been made and have also verified that the output from the new version is similar to the previous.

The addition of checks to ensure UTF-8 encoding in the `data1.csv` file also adds an extra layer of confidence that the inputs are as intended. I saved the input file `SAS_indata_TampaBayTP-03-22-22.csv` in the `UserTutorialDynamic/data` folder using ANSI encoding and the model run exited on data input with the appropriate messages indicating incorrect encoding. However, the encoding was flagged as ISO-8859-1 encoding,

which I suppose is similar, but not the same. This is a minor issue that I'm just mentioning in case more accurate identification is needed.

It also appears that some of the previous code styling suggestions were incorporated, particularly in the use of spacing for `if/else` statements, the assignment operators, and mathematical operations. These are all good changes that have improved legibility of the code. There were also a number of minor changes I suggested that were incorporated, e.g., base functions are now used in place of generic functions that called the base functions (e.g., lines 40, 41, 54, and 55 in `hydseq.R`).

Code testing in the `tests/testthat` folder has also been implemented following previous suggestions from myself and an earlier review. I know this is a difficult package to develop testing, but it might be worth seeing what kind of test coverage is provided, i.e., what percentage of the code is actually tested. This can be done using the `covr` package, which provides a function `package_coverage()` that will return a summary of the test coverage for all functions in the package. I ran this on the local version of the package and it showed only 5% coverage. Writing tests is tedious and even more challenging to write good tests. It is not a requirement for an R package, but generally good practice, so I encourage additional tests to be written in future versions. Based on the issue posted in repository, I gather these tests are primarily for functionality that was previously provided by .Rmd files, so I am not surprised the coverage is low but still useful for the additions.

One new potential issue I found is the declaration of the colors for some of the graphics. For example, on line 713 in `sparrow_control.R`, there are colors named as `"light blue"`, `"dark green"`, and `"dark red"`. These are not standard color names in R, although they seem to be working for some of the graphics on the Shiny app. This may cause issues for other graphing devices or versions of R, so I suggest changing them to the standard naming convention that does not use spaces. This should be easy to update using "Find" and "Replace All" in RStudio.

Domain review

Because none of the core concepts/science of *RSPARROW* were modified for this update, the domain review covers the changes to the documentation in the file `RSPARROW_docV2.1.pdf` in the folder `inst/doc` as identified in a documentation update file provided via email and available here. The following are minor comments on these updates.

- The citation for the new reference Miller et al. 2024 has a typo in the title - "[emporal" should be "Temporal".
- Section 8.1, perhaps include a URL for CRAN and RTools to direct users to the software downloads should they wish to install on their own.
- Fig. 55, although this in the text, perhaps include the full path to `funcTypes.csv` in the caption. Also, it doesn't really seem necessary to include Figs. 56-59 since these are just continuations of the info in Fig. 55, especially since not all twelve function category

columns are present. The intent is just to show what the file looks like, not replace actually viewing the file.

- Section 8.9, bullet 5, this seems like a relatively simple future update that would potentially expand the user base (e.g., many are using IOS). It is also standard practice to check R package builds on different operating systems and including this interoperability could facilitate future work to make *RSPARROW* more portable (e.g., on CRAN or some other binary distribution system).

All other items in the documentation update file were verified and no issues were found.