tbeptools: An R package for synthesizing estuarine data for environmental research

Marcus W. Beck[*,1]**, Meagan N. Schrandt**[2]**, Michael R. Wessel**[3]**, Edward T. Sherwood**[1]**, Gary E. Raulerson**[1]**, Adhokshaja Achar Budihal Prasad**[4]**, and Benjamin D. Best**[5]

1 Tampa Bay Estuary Program, St. Petersburg, Florida, USA **2** Fish and Wildlife Research Institute, Florida Fish and Wildlife Conservation Commission, St. Petersburg, Florida, USA **3** Janicki Environmental, Inc., St. Petersburg, Florida, USA **4** University of South Florida, Tampa, Florida, USA **5** EcoQuants, LLC, Santa Barbara, California, USA

Summary

Many environmental programs report on the status and trends of natural resources to inform management decisions for protecting or restoring environmental condition. The National Estuary Program (NEP) in the United States is one example of a resource management institution focused on "estuaries of national significance" that provides place-based solutions to managing coastal resources. There are 28 NEPs in the United States, each with similar but location-specific programmatic goals to address environmental challenges related to water quality, alteration of hydrologic flows, invasive species, climate change, declines in fish and wildlife populations, pathogens and other contaminants, and stormwater management. A critical need of each NEP is the synthesis of data from disparate sources that can inform management response to address these environmental challenges.

The Tampa Bay Estuary Program (TBEP) in Florida, USA is responsible for developing and implementing a place-based plan to sustain historical and future progress in the restoration of Tampa Bay (N. O'Hara, Shafer Consulting, Inc., 2017). The needs of TBEP for reporting on indicators of environmental condition are similar to other environmental organizations. Multiple local and regional partners collect data that are used for different reporting products. Without data synthesis tools that are transparent, accessible, and reproducible, NEP staff and colleagues waste time and resources compiling information by hand. The `tbeptools` R software package can be used for routine development of reporting products, allowing for more efficient use of limited resources and a more effective approach to communicate research to environmental decision-makers. Functions in `tbeptools` also support the creation of content for interactive, online dashboards that can facilitate more informed decisions without requiring an intimate understanding of the R programming language or the methods for analysis.

The `tbeptools` package also addresses challenges associated with data retrieval and assessment of environmental data relative to important policy or management targets. For each environmental indicator, functions are included to import required data directly from sources, removing the need to manually obtain information prior to reporting. These functions are integrated into summary report cards that are generated automatically through continuous integration services (i.e., GitHub Actions) that free the analyst from external downloads, analysis, and copying of results that can introduce errors in reporting. Similar packages provide seamless access to database services (e.g., `dataRetrieval`, De Cicco, Lorenz, Hirsch, Watkins, & Johnson, 2021), but few packages link these data sources

---

[*]Corresponding author

directly to analysis and reporting as in `tbeptools` (but see `wqindex`, Thorley, Millar, & Teucher, 2018). Management targets and regulatory thresholds based on summary assessments of data, either direct from sources or included as supplementary data in the package, are also hard-coded into the functions.

Statement of need

The `tbeptools` R package was developed to automate data synthesis and analysis for many of the environmental indicators for Tampa Bay, with more general application to commonly available datasets for estuaries. The functions in the package were developed to extract methods from existing technical documents and to make them available in an open source programming environment. By making these tools available as an R package, routine assessments are now accomplished more quickly and other researchers can use the tools to develop more specific analysis pipelines.

Most of the NEPs do not have analysis software to operationalize data import, analysis, and plotting for reporting. Recently, a similar software package, `peptools` (Marcus Beck, 2021), was developed for the Peconic Estuary Partnership (New York, USA) using many of the functions in `tbeptools` to develop reporting products for a new water quality monitoring program. This successful technology transfer demonstrates the added value of presenting these methods in an open source environment available for discovery and reuse by others. We expect other NEPs to begin using these tools as their application becomes more widespread among estuarine researchers.

Beyond the NEPs, `tbeptools` is an effective example of an R package for implementing technical methods in existing literature and reports that can be used to support environmental monitoring and assessment needs for science-based decisions. To this end, the `tbeptools` package was also created to support the development of online dashboards created in R Shiny (Chang et al., 2021). Dashboards are powerful tools to increase accessibility for end users to engage with scientific products without the need to understand technical details in their creation. However, providing the underlying methods as source code in an R package increases transparency and reproducibility of reporting products if users require a more detailed understanding of how the content was created. Currently, the `tbeptools` package supports dashboards created by TBEP for the assessment of water quality (Figure 1, M. W. Beck, 2020a), seagrasses (M. W. Beck, 2020b), nekton communities (M. W. Beck, 2020c), and tidal creeks (M. W. Beck & Wessel, 2020). Resource management agencies or similar institutions could follow this approach to facilitate development of front-end products for more informed decision-making.

Example usage

The function names were chosen with a typical analysis workflow in mind, where functions are available to `read` data from a source (typically from an online repository or stable URL), `anlz` to analyze the imported data using methods in existing technical documents or published papers, and to `show` the results as a summary graphic for use by environmental managers. The functions are used to report on water quality (M. Beck, Burke, & Raulerson, 2021), fisheries (Schrandt, MacDonald, Sherwood, & Beck, 2021), benthic condition (D.J. Karlen, T. Dix, B.K. Goetting, S.E. Markham, K.Campbell, J. Jernigan, J.Christian, K. Martinez, A. Chacour, 2020), tidal creeks (Wessel et al., 2021), and seagrass transect data (Sherwood, Greening, Johansson, Kaufman, & Raulerson, 2017). The vignettes for the package are topically organized to describe the functions that apply to each of the indicators.

The following example demonstrates use of a subset of the functions for water quality data to read a file from the Hillsborough County Environmental Protection Commission long-
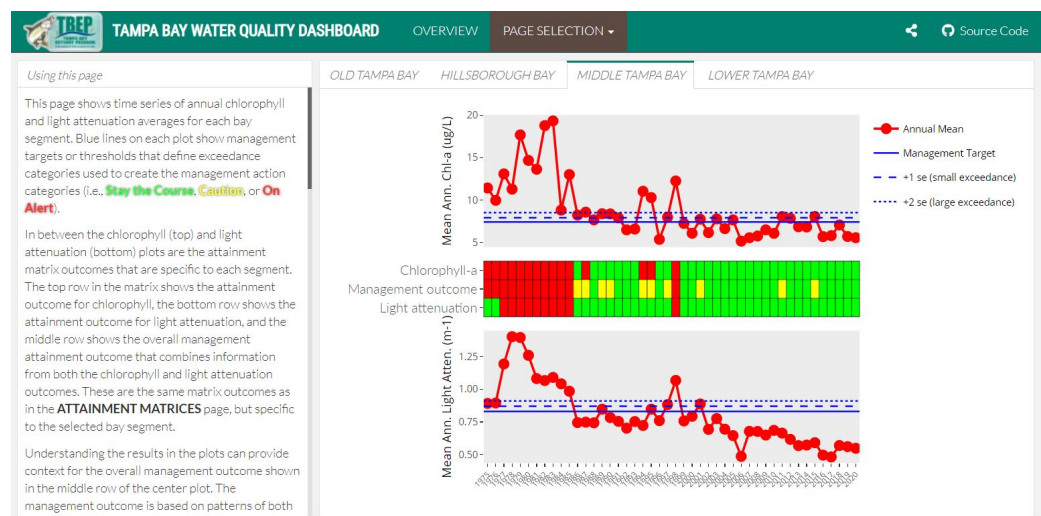
Figure 1: The TBEP water quality dashboard, demonstrating use of the `tbeptools` R package to generate summary plots for specific bay segments.

term monitoring dataset (available from https://www.tampabay.wateratlas.usf.edu/), analyze monthly and annual averages by major bay segments of Tampa Bay, and plot an annual time series for one of the bay segments.

```r
# load the package
library(tbeptools)

# read current data
wqdat <- read_importwq(xlsx = 'wqdata.xlsx', download_latest = TRUE)
wqdat
```

```
## # A tibble: 26,611 x 22
##   bay_segment epchc_station SampleTime            yr    mo
##   <chr>               <dbl> <dttm>              <dbl> <dbl>
## 1 HB                      6 2021-06-08 10:59:00  2021     6
## 2 HB                      7 2021-06-08 11:13:00  2021     6
## 3 HB                      8 2021-06-08 14:15:00  2021     6
## 4 MTB                     9 2021-06-08 13:14:00  2021     6
## 5 MTB                    11 2021-06-08 11:30:00  2021     6
## # ... with 26,606 more rows, and 17 more variables:
## #   Latitude <dbl>, Longitude <dbl>, Total_Depth_m <dbl>,
## #   Sample_Depth_m <dbl>, tn <dbl>, tn_q <chr>, sd_m <dbl>,
## #   sd_raw_m <dbl>, sd_q <chr>, chla <dbl>, chla_q <chr>,
## #   Sal_Top_ppth <dbl>, Sal_Mid_ppth <dbl>,
## #   Sal_Bottom_ppth <dbl>, Temp_Water_Top_degC <dbl>,
## #   Temp_Water_Mid_degC <dbl>, ...
```

```r
# analyze monthly and annual means by bay segment
avedat <- anlz_avedat(wqdat)
avedat
```

```
## $ann
## # A tibble: 584 x 4
##      yr bay_segment var          val
##   <dbl> <chr>       <chr>      <dbl>
## 1  1974 HB          mean_chla   22.4
```

, (). tbeptools: An R package for synthesizing estuarine data for environmental research. *Journal of Open Source Software*, (), . https://doi.org

```
## 2  1974 LTB        mean_chla  4.24
## 3  1974 MTB        mean_chla  9.66
## 4  1974 OTB        mean_chla 10.2
## 5  1975 HB         mean_chla 27.9
## # ... with 579 more rows
##
## $mos
## # A tibble: 4,484 x 5
##   bay_segment    yr    mo var          val
##   <chr>        <dbl> <dbl> <chr>      <dbl>
## 1 HB            1974     1 mean_chla 36.2
## 2 LTB           1974     1 mean_chla  1.75
## 3 MTB           1974     1 mean_chla 11.5
## 4 OTB           1974     1 mean_chla  4.4
## 5 HB            1974     2 mean_chla 42.4
## # ... with 4,479 more rows
```

```
# show annual time series of chlorophyll for Hillsborough bay segment
show_thrplot(wqdat, bay_segment = 'HB', yrrng = c(1975, 2020))
```
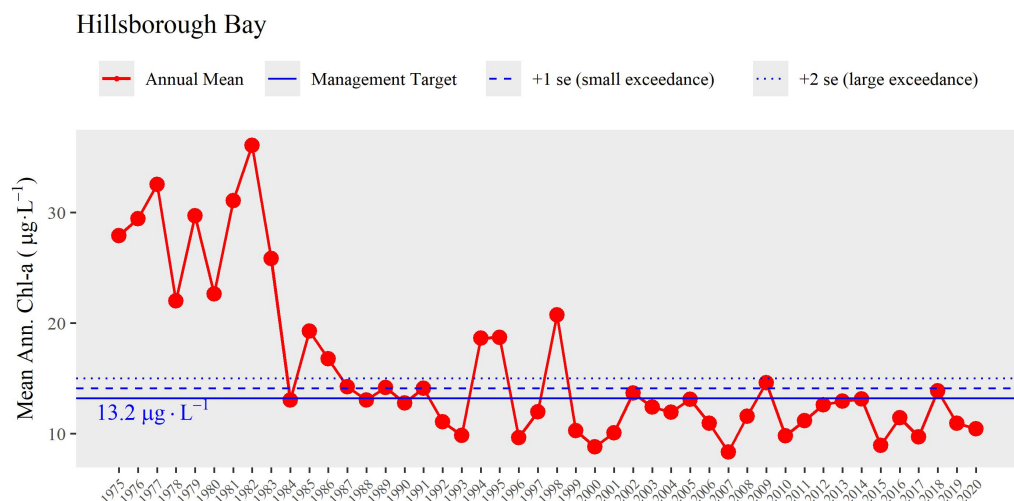

Acknowledgements

We acknowledge our many local and regional partners for their continuing collaborative efforts in working towards a healthy Tampa Bay, in particular the Tampa Bay Nitrogen Management Consortium. The `tbeptools` software would not be possible without data provided by our partners. We also thank two reviewers for providing useful feedback that improved the manuscript.

References

Beck, Marcus. (2021). Peptools: Analysis tools for importing, wrangling, and summarizing Suffolk County water quality data. R package version 0.0.0.9004. https://github.com/tbep-tech/peptools.

Beck, M., Burke, M., & Raulerson, G. (2021). *2020 Tampa Bay Water Quality Assessment* (No. 05-21) (p. 2). St. Petersburg, Florida: Tampa Bay Estuary Program.

Beck, M. W. (2020a). *Tbep-tech/wq-dash: v1.0.* Zenodo. doi:10.5281/zenodo.3648664

Beck, M. W. (2020b). *Tbep-tech/seagrasstransect-dash: v1.0*. Zenodo. doi:10.5281/zenodo.4319936

Beck, M. W. (2020c). *Tbep-tech/nekton-dash: v1.0*. Zenodo. doi:10.5281/zenodo.3939236

Beck, M. W., & Wessel, M. R. (2020). *Tbep-tech/tidalcreek-dash: v1.0*. Zenodo. doi:10.5281/zenodo.3701560

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., et al. (2021). *Shiny: Web application framework for R*. Retrieved from https://CRAN.R-project.org/package=shiny

D.J. Karlen, T. Dix, B.K. Goetting, S.E. Markham, K.Campbell, J. Jernigan, J.Christian, K. Martinez, A. Chacour. (2020). *Twenty-five-year Trends in the Benthic Community and Sediment Quality of Tampa Bay 1993-2017* (No. 06-20) (p. 251). St. Petersburg, Florida: Tampa Bay Estuary Program.

De Cicco, L. A., Lorenz, D., Hirsch, R. M., Watkins, W., & Johnson, M. (2021). *dataRetrieval: R packages for discovering and retrieving water data available from U.S. Federal hydrologic web services*. Reston, VA: U.S. Geological Survey; U.S. Geological Survey. doi:10.5066/P9X4L3GE

N. O'Hara, Shafer Consulting, Inc. (2017). *Charting the Course: The Comprehensive Conservation and Management Plan for Tampa Bay* (No. 10-17). St. Petersburg, Florida: Tampa Bay Estuary Program. Retrieved from https://drive.google.com/file/d/1HgnP5dC2LQqffc6Bq6klEuSUcNC2WRiZ/view?usp=drivesdk

Schrandt, M. N., MacDonald, T. C., Sherwood, E. T., & Beck, M. W. (2021). A multimetric nekton index for monitoring, managing and communicating ecosystem health status in an urbanized gulf of mexico estuary. *Ecological Indicators*, *123*, 107310. doi:10.1016/j.ecolind.2020.107310

Sherwood, E., Greening, H., Johansson, J. O. R., Kaufman, K., & Raulerson, G. (2017). Tampa Bay (Florida, USA): Documenting seagrass recovery since the 1980's and reviewing the benefits. *Southeastern Geographer*, *57*(3), 294–319. doi:10.1353/sgo.2017.0026

Thorley, J., Millar, C., & Teucher, A. (2018). *Wqindex: Water quality index calculation for British Columbia*. Retrieved from https://github.com/bcgov/wqindex

Wessel, M. R., Leverone, J. R., Beck, M. W., Sherwood, E. T., Hecker, J., West, S., & Janicki, A. (2021). Developing a water quality assessment framework for southwest florida tidal creeks. *Estuaries and Coasts*, *In press*. doi:10.1007/s12237-021-00974-7