hatchR: A toolset to predict when fish hatch and emerge

Morgan M. Sparks[1,4], Eli A. Felts[2], Allison G. Swartz[3], Bryan M. Maitland[1]

[1]Rocky Mountain Research Station, U.S. Forest Service, Boise, ID, USA

[2]Clearwater Solutions, Lenore, ID, USA

[3]College of Forestry, Oregon State University, Corvallis, OR, USA

[4]Corresponding Author: Morgan M. Sparks, Morgan.Sparks@usda.gov

keywords: developmental phenology, effective value model, R, Shiny

Abstract

Understanding the timing of key life history events is essential for effective fish conservation and management. Traditionally, predicting hatch and emergence timing in wild fish populations was challenging due to the reliance on average incubation temperature as a primary model parameter, which is often difficult in to obtain in natural settings. Recent advancements have refined these models, enabling their application in wild environments using spawning dates and daily water temperature records. However, their broader use remains constrained by a lack of parameterizations for many species, with most applications focused on Salmonids. Here we introduce **hatchR**, a software ecosystem designed to predict hatch and emergence for a wide range of wild fishes. **hatchR** offers users access to established phenological models and the flexibility to incorporate custom parameterizations using external datasets. The software is available in two formats: an open-source R package for advanced customization and an HTML-based graphical user interface for those unfamiliar with scripting. To illustrate its utility, we present two case studies demonstrating its application in research and management. By expanding access to predictive modeling tools, **hatchR** has the potential to advance studies of fish early life history and support conservation efforts across diverse species.

Introduction

As poikilothermic organisms, fish development and growth are closely tied to ambient environment. This strong relationship has enabled researchers to generate statistical models that predict developmental phenology with high accuracy. Historically, these models were formulated in aquaculture settings under the assumption of constant temperature throughout development (Alderdice & Velsen, 1978; Beacham & Murray, 1990; McPhail & Murray, 1979), limiting their applicability to wild populations. However, Sparks et al. (2019) reformulated this approach into an "Effective Value model", which instead uses daily average temperature after spawning, predicting hatch or emergence when cumulative effective values reach a threshold of one.

This effective value approach has since been widely applied to Salmonids, for which aquaculture-derived parameterizations were readily available. For example, Pacific Salmon (*Oncorhynchus* spp.) models developed by Beacham & Murray (1990) have been applied across various species and populations (Adelfio, Wondzell, Mantua, & Reeves, 2019, 2024; Kaylor et al., 2021), while Bull Trout (*Salvelinus confluentus*) models from McPhail & Murray (1979) were extended by Austin, Essington, & Quinn (2019). Despite its growing adoption, applications of the effective value model remain largely confined to Salmonids, likely due to the availability of existing parameterizations and the commercial and recreational importance of these species.

To extend these modeling capabilities beyond Salmonids and facilitate broader applications, we developed **hatchR**, a software ecosystem designed to predict hatch and emergence timing for wild fish populations. **hatchR** enables users to input standard raw or summarized water temperature datasets commonly collected in field settings, conduct basic data validation, and apply built-in parameterizations such as those from Beacham & Murray (1990) or Sparks, Westley, Falke, & Quinn (2017). Users can also develop custom models using their own or published temperature and phenological data within the effective value framework.

To maximize accessibility, **hatchR** is available in two formats. The first is a R package, **hatchR**, distributed via CRAN ("R," n.d.), providing advanced customization and automation for analyzing multiple variables, such as phenology type, spawn timing, or thermal regimes. Comprehensive documentation is available on the **hatchR** website (https://bmait101.github.io/hatchR/). The second is a Shiny-based web application (Chang et al., 2024), offering a graphical user interface for those unfamiliar with R, balancing ease of use with much of the R package's core functionality. Below, we provide and overview of **hatchR** and present case studies demonstrating its application in research and management.

Package Overview

hatchR is designed primarily as a tool for predicting fish phenology. To maintain focus on this core function, we provide minimal built-in data validation and visualization tools, as users are expected to understand and check their own data. Given the diversity of potential data types, it is impractical to implement comprehensive validation checks. However, we include basic data-checking and summarization functions (`check_continuous()`, `summarize_temp()`) and limited built-in visualization capabilities (`plot_check_temp()`, `plot_phenology()`). Intuitive functions are provided for users to apply models—either existing models from the literature using the `model_select()` function or fitting custom functions from data using the `fit_model()` function. Users can then apply these models to water temperature data (*e.g.,* from a HOBO temperature logger) to predict when hatching phenology will occur. This is accomplished with the `predict_phenology()` function. The R package provides example workflows for customizing plots from model output, while the Shiny application includes a default output plot and an option to download results for external visualizations. For a high-level overview of **hatchR**'s applications, see Figure 1. Additional details on key functions and workflows–particularly for automating phenology predictions across multiple variables–are available in articles hosted on the software's webpage.

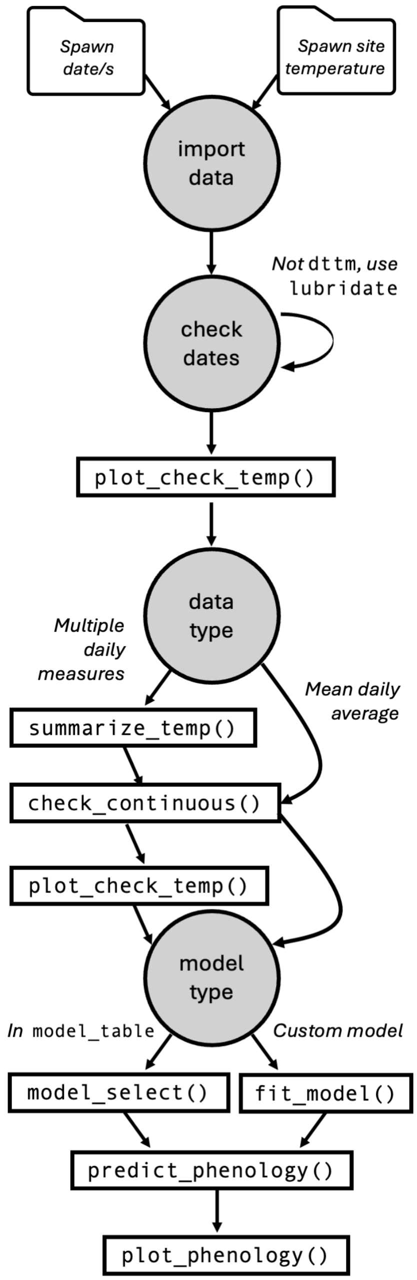

Figure 1: hatchR workflow. Data inputs are represented by folders, data processes by filled circles, hatchR functions as plain text rectangles, and decision choices as italicized text.

Effective value models

Effective value models were introduced by Sparks et al. (2019) to predict developmental timing in wild populations, initially for Sockeye Salmon (*O. nerka*). Their development was necessitated by limitations in traditional models, such as those in Beacham & Murray (1990), which relied on average incubation temperature over the full developmental period. In wild settings, estimating this average temperature was impracticable since hatch timing was unknown, even when spawning dates were recorded. To overcome this challenge, Sparks et al. (2019) reformulated model 2 from Beacham & Murray (1990) by taking its reciprocal and assigning an *effective value* to each day of development based on the daily average temperature. This approach allowed for cumulative tracking of developmental progress, enabling hatch and emergence predictions without requiring prior knowledge of incubation temperatures.

The model follows the general format of:

$$\mathrm{E_i} = \frac{1}{\exp(\log_e(a) - \log_e(T_i - b)}$$

where E_i is the effective value and T_i the temperature for day i, and a and b are model parameterization estimates (i.e. species- or model-specific constants). A fish hatches or emerges when the cumulative sum of effective values reaches one:

$$\sum_{i=1}^{n} E_i = 1$$

This framework as the foundation for phenological models in **hatchR**. The package includes a predefined `model_table` containing established parameterizations, including those from Beacham & Murray (1990), Sparks et al. (2017), and Austin et al. (2019) (who extended McPhail & Murray (1979)). While `model_table` incorporated more complex models from Beacham & Murray (1990), users can also fit custom models using the `fit_model()` function. This flexibility allows for the incorporation of new parameterizations as they are developed, expanding the utility of **hatchR** beyond Salmonids.

Data format

Water temperature datasets collected in the field typically fall into two categories: 1) summarized daily data, where mean daily temperatures are pre-computed, or 2) raw high-frequency data, such as those recorded by HOBO TidbiT loggers, which require summary into mean daily temperatures before use. Additionally, new statistical models, such as Siegel, Fullerton, FitzGerald, Holzer, & Jordan (2023), could also be implemented into this framework.

hatchR assumes input data consists if at least two required columns: a date column indicating the date (and optionally time) of each temperature measurement, and a temperature column providing the corresponding temperature measurement (in °C). Other columns may be present, but columns names should not include spaces. Data should follow the format outlined in Table 1.

Table 1: Example temperature data for use in hatchR.

date	temperature
2000-01-01	2.51
...	...
2000-07-01	16.32
...	...
2000-12-31	3.13

Since **hatchR** does not automatically handle missing data, users must check for gaps or errors before running analyses. The package will function with missing values, but gaps in the dataset may affect predictions. **hatchR** supports temperatures as low as 0 °C, though such values yield extremely small effective values, potentially extending hatch or emergence timing to a year or more. Users should critically assess whether such data align with biological expectations.

For R users, **hatchR** can import data in any format, provided it is converted into a `data.frame` or `tibble`, where each row represents a single temperature record. The Shiny application requires data to be uploaded as a .csv (comma separated values) file, which can easily be exported from spreadsheet software such as Microsoft Excel or Google Sheets.

Checking Data

hatchR is designed to analyze daily average temperatures. While high-frequency data (e.g., from HOBO loggers) can be used, it must be summarized into daily averages. **hatchR** provides built-in functionality for this summarization in R but requires pre-summarized data for use in the Shiny app.

To help users identify potential issues, **hatchR** includes basic data checking functions that highlight outliers or missing values both visually and programatically. These checks ensure data integrity before model application.

We demonstrate the utility three functions: `summarize_temp()`, `plot_check_temp()`, and `check_continuous()`– using a simulated year-long data set (`year_sim`). This dataset contains temperature readings taken every thirty minutes, and its structure (dimensions and first six rows) is shown below.

```r
#year_sim data dimensions (rows x columns)
dim(year_sim)
```

```
## [1] 17568     2
```

```r
#fist 6 rows of year_sim
head(year_sim)
```

```
##                  date      temp
## 1 2000-07-01 00:00:00  8.318573
## 2 2000-07-01 00:29:55  9.309468
## 3 2000-07-01 00:59:50 14.676125
## 4 2000-07-01 01:29:45 10.211525
## 5 2000-07-01 01:59:40 10.387863
## 6 2000-07-01 02:29:35 15.145195
```

First, we recommend using `plot_check_temp()` to visually inspect imported data for outliers or unusual values (Figure 2).

```r
plot_check_temp(data = year_sim,
                dates = date,
                temperature = temp,
                temp_min = 0, # temp_min and max lines are
                temp_max = 25) # user customizable
```

In this case, no obvious outliers are present, but since each day contains 48 records, the data must be summarized to daily mean temperature using `summarize_temp()`. After summarization, `check_continuous()` should be used to identify any missing days. We also suggest running `plot_check_temp()` again on the summarized data to verify its integrity, though we omit the resulting plot here for space efficiency.

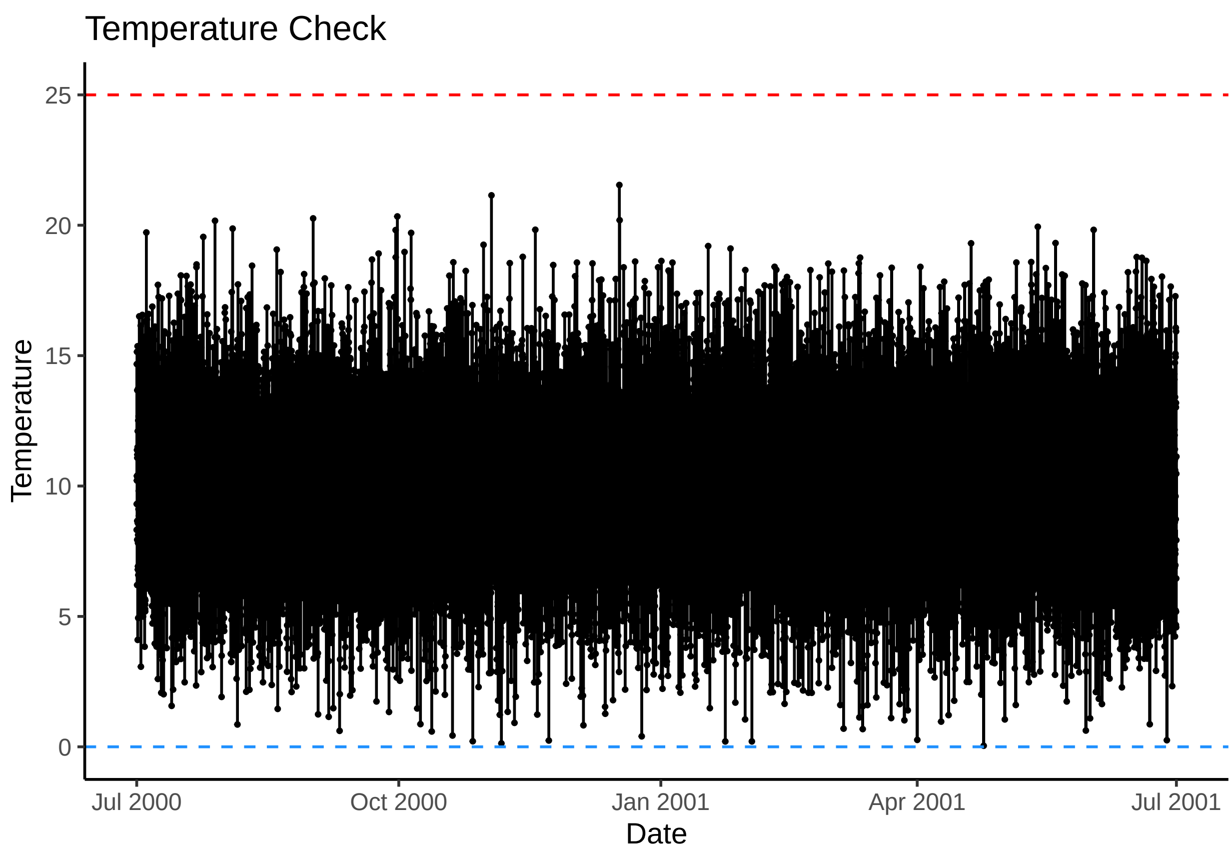

Figure 2: Output of hatchR function `plot_check_temp()`, which is used as a visual data check on the raw `year_sim` data set. Users can set custom thresholds for minimum and maximum temperatures (dashed lines).

```r
# summarize
year_sim_summ <- summarize_temp(data = year_sim,
                                dates = date,
                                temperature = temp)

# now a year's worth of single-day data
# 365 days x two columns
dim(year_sim_summ)
```

```
## [1] 365   2
```

```r
# check continuous (no errors)
check_continuous(data = year_sim_summ,
                 dates = date)
```

```
## i No breaks were found. All clear!
```

```r
# we can demonstrate an error by removing Oct. 8 (100th day)
check_continuous(data = year_sim_summ[-100,],
                 dates = date)
```

```
## Warning: ! Data not continuous
## i Breaks found at rows:
## i 100
```

Model Selection

Users can select from existing Salmonid models in `model_table` or generate custom models using `fit_model()` in both the R and Shiny deployments of **hatchR**. The models in `model_table` are included because their parameterizations are well-documented in the literature, though they are currently limited to Pacific Salmon and Bull Trout (see Quinn (2018) pg. 183, for additional Salmonid models). To ensure reliability, we restrict `model_table` to well vetted models with experimental ranges spanning 2-17 °C.

Custom models, by contrast, often have narrower parameterization ranges. To prevent misapplication, we exclude other model parameterization from `model_table`, requiring users to carefully asses whether their parameterized models are appropriate for the temperature ranges in their datasets.

To expand the applicability of the effective value approach beyond Salmonids, **hatchR** includes a `fit_model()` function, which is species-agnostic as long as development follows a power law relationship with temperature. The function takes two input vectors: average incubation temperature (°C) and the number of days to a given phenological event. A model is then fit to the data using `stats::nls()`, which performs nonlinear least squares regression to estimate the parameters $log_e a$ and b. Because the optimization process in `nls()` is sensitive to initial parameter values, `fit_model()` first fits a linear model to the log-transformed data to provide initial parameter estimates, which are then used to fit the nonlinear model.

This approach allows users to generate models tailored to non-Salmonid species, provided they have experimental or field data linking development to temperature. However, users should be mindful of several factors, such as extrapolations risks (models may not generalize beyond the temperature range for which they were parameterized) and species-specific variation (genetic differences among populations may affect developmental responses). Future expansions of **hatchR** could incorporate additional vetted parameterizations for other taxa, such as non-Salmonid fishes, amphibians, invertebrates, provided sufficient validation in the literature.

Fitting models for other fishes

We demonstrate how the `fit_model()` function may be used to create custom parameterizations for species beyond the Salmonids included in `model_table`. To showcase its utility, we provide parameterizations for three warm-water species: Smallmouth Bass (*Micropterus dolomieu*) (Webster, 1948), Channel Catfish (*Ictalurus punctatus*) (Small & Bates, 2001), and Lake Sturgeon (*Acipenser fulvescens*) (Smith & King, 2005). These species were selected due to their common use in aquaculture and sport fisheries, illustrating the broad applicability of the effective values approach. For concision, we present parameterization for Smallmouth Bass here, while the full implementation details for all species are available in the `paper.Rmd` on the GitHub project repository.

To demonstrate parameterizations, we generate a simulated thermal regime featuring an ascending thermograph with a mean temperature of 16 °C (again, available in paper.Rmd). Using this dataset, we apply the parameterized models for each species to predict hatch timing and total developmental duration (Figure 3). This example highlights the flexibility of **hatchR** for accommodating diverse fish species and environmental conditions, making it a valuable tool for researchers and managers working outside of Salmonid systems.

```r
### smallmouth mod
smallmouth <- matrix(NA, 10, 2) |> data.frame()
colnames(smallmouth) <- c("hours", "temp_F")
smallmouth$hours <- c(52, 54, 70, 78, 90, 98, 150, 167, 238, 234)
smallmouth$temp_F <- c(77, 75, 71, 70, 67, 65, 60, 59, 55, 55)

# change °F to °C and hours to days
smallmouth <- smallmouth |>
  mutate(days = ceiling(hours/24),
         temp_C = (temp_F -32) * (5/9))

# model object for smallmouth bass
smb_mod <- fit_model(temp = smallmouth$temp_C,
                     days = smallmouth$days,
                     species = "smb",
                     development_type = "hatch")
```

Note the R^2 fit from the models below. You can see they generally preform well and are close to values from model 2 of Beacham & Murray (1990), which fall between 0.95-0.99 range.

```r
#model fits
smb_mod$r_squared; cat_mod$r_squared; sturgeon_mod$r_squared
```

```
## [1] 0.9868067

## [1] 0.9433598

## [1] 0.9217358
```

Predicting Phenology and Output

To illustrate model selection and phenology prediction, we will replicate a portion of the analysis from Sparks et al. (2019) using the `woody_island` dataset included with **hatchR**. Specifically, we predict both hatch and emergence timing for Sockeye Salmon at Woody Island in 1990.

First, we obtain model expression for both hatch and emergence using `model_select()`, which retrieves the appropriate parameterizations from `model_table`:

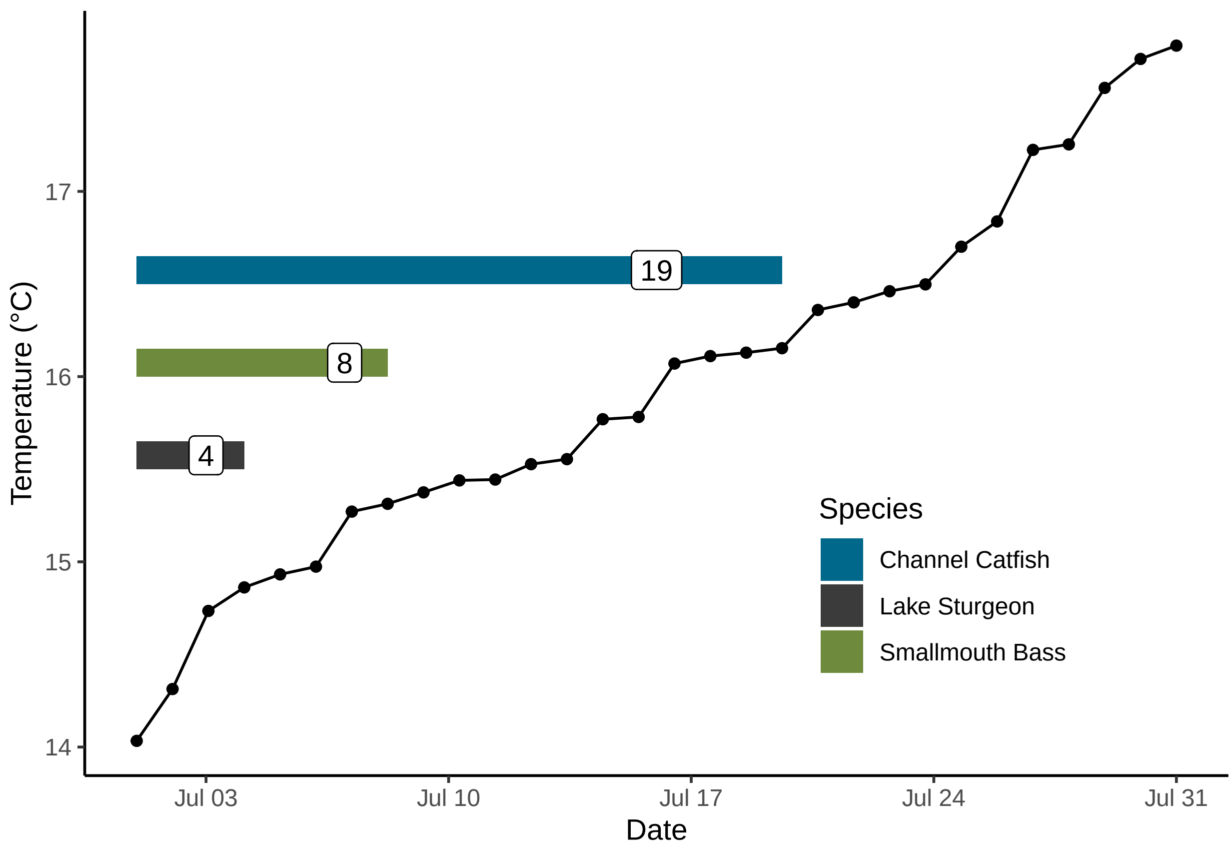

Figure 3: Predicted days to hatch for three warmwater species with custom parameterizations using a random thermal regime with an ascending thermograph with a mean temperature of 16 °C over 30 days.

```r
#model_select() to get sockeye model
sockeye_hatch_mod <- model_select(
  author = "Beacham and Murray 1990",
  species = "sockeye",
  model = 2,
  development_type = "hatch"
  )
```

These model expressions are then implemented in `predict_phenology()` to estimate the days to hatch and development:

```r
#predict_phenology() with sockeye model and woody_island temperature data
WI_hatch <- predict_phenology(
  data = woody_island,
  dates = date,
  temperature = temp_c,
  spawn.date = "1990-08-18", #notice the character string for spawn date
  model = sockeye_hatch_mod
  )
```

The returned object provides key outputs, including days to hatch and the full developmental period, allowing us to assess phenological patterns under the recorded thermal conditions:

```r
WI_hatch$days_to_develop
```

```
## [1] 74
```

```r
WI_hatch$dev.period
```

```
##        start       stop
## 1 1990-08-18 1990-10-30
```

Understanding your results

The output from `predict_phenology()` contains multiple elements in a list, which can be accessed using the $ operator. Each component provides different insight into the predicted phenology:

```r
summary(WI_hatch)
```

```
##                 Length Class       Mode
## days_to_develop 1      -none-      numeric
## dev.period      2      data.frame  list
## ef_table        5      tbl_df      list
## model_specs     5      spec_tbl_df list
```

`WI_hatch$days_to_develop` – Returns the predicted number of days required for development.

`WI_hatch$dev.period` – A 1x2 dataframe containing the spawning date (as input via `predict_phenology(spawn.date = ...)`) and predicted development completion date.

`WI_hatch$ef_table` – An n x 5 tibble (n = number of days to hatch or emerge), containing a row index, the date, each day's temperature and effective value, and the cumulative sum of the effective values. This table serves as a foundation for users to create custom visualizations beyond the built-in functionality discuss below.

`WI_hatch$model_specs` – Provides details about the model used for prediction, including whether it was retrieved from `model_select()` or generated using `fit_model()`. Most importantly, it contains the model expression (*i.e.*, the formula) used for phenology predictions.

Plotting phenology

hatchR includes a built in function, `plot_phenology()`, for visualizing phenology predictions (Figure 4). This function generates plots with three specific components: 1) the temperature regime over the prediction period, 2) the cumulative sum of effective values, and 3) the effective value for each day within the prediction span. Be default, `plot_phenology()` produces a comprehensive figure that includes all three elements with corresponding labels and titles. However, users can customize the output to focus on specific aspects of interest, allowing for tailored visual representations of their results.

This function provides a quick and effective way to interpret model outputs, facilitating comparisons between temperature regimes or species-specific phenological responses.

```r
plot_phenology(WI_hatch)
```

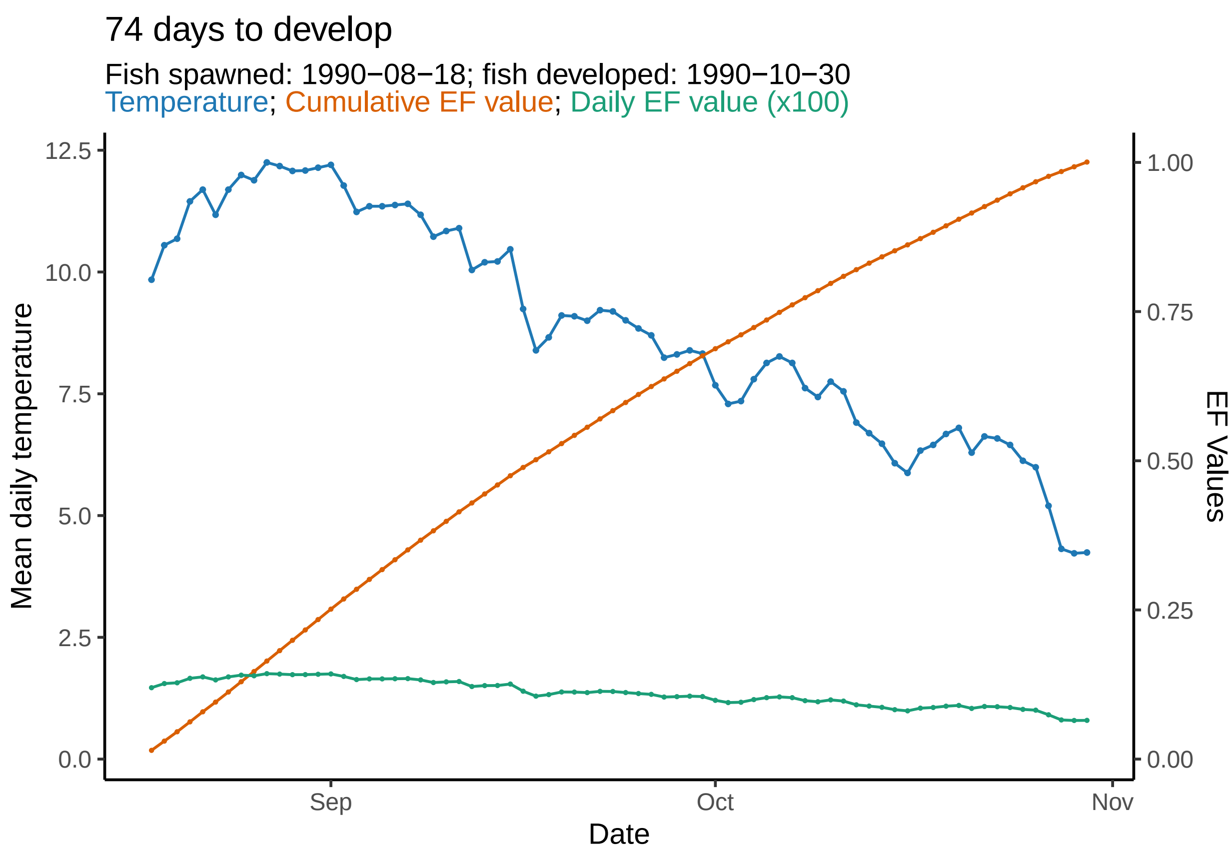

Figure 4: Output of `plot_phenolgy()` function using predicted hatch time from `woody_island` data set.

Case Study 1

Predicting Emergence Timing for Management Actions

A common management application of developmental phenology is assessing whether fish will be free-moving before a scheduled management action, such as stream section access for grazing or road work. For instance, will fish have emerged from redds before construction begins, reducing potential disturbance?

In this scenario, we consider road work near the upper portion of Crooked River in the Boise River watershed, home to a key Bull Trout (*Salvelinus confluentus*) population. Bull Trout, a federally threatened species under the Endangered Species Act (Nolfi, Melbihess, Fisher, & Ellis, 2024), are particularly sensitive to sediment disturbance. The Forest Service Fisheries Biologist overseeing the project wants to determine whether Bull Trout fry will likely be out of the gravel and free-swimming by June 1st. In this system, Bull Trout typically complete spawning by the end of September, so we consider the latest possible spawn date: September 30th.

To demonstrate this case study, we use the **hatchR** graphical user interface available at https://elifelts.shinyapps.io/hatchR_demo/. Users begin by uploading their temperature dataset through the `Import Data` window, selecting their file, specifying the appropriate temperature and date columns, and providing the date format (e.g., year-month-day or day-month-year). For this example, we use the `crooked_river` dataset, included in the Shiny app as a demo dataset. It can also be accessed directly at https://github.com/bmait101/hatchR/blob/master/extdata/crooked_river.csv/.

Once uploaded, **hatchR** automatically generates a visual data check using `plot_check_temp()`. After confirming data integrity, users navigate to the `Model Phenology` window. For this case study, we use the pre-loaded Bull Trout parameterization from Austin et al. (2019), selecting the `Existing` model option via dropdown menus. The user then chooses multiple spawn dates using an interactive calendar. Here, we focus on September 30th (in the 2014 spawn year) as outlined in our management scenario.

Following model selection, **hatchR** outputs results in two key locations: the `Phenology Summaries` tab, which provides a table with predictions for each spawn date, and the `Timeline Plot` tab, which shows the corresponding visualization of development timing. Both the prediction table and plot can be downloaded directly from their respective tabs.

The process is demonstrated in the included ***supplementary video file???***, and a more detailed interface walk through is available on the **hatchR** Shiny website.

In this example, the model predicts that the last Bull Trout will emerge before the June 1st road work target date. This suggests that the Fisheries Biologist can confidently approve the work in the area without concern for sediment disturbance impacting fish developing in the gravel. This type of predictive modeling helps managers make informed, science-based decisions that balance conservation priorities with land-use activities.

Case Study 2

Large Scale Predictions of Bull Trout Development Timing

For the second case study, we demonstrate a more complex, large-scale application of **hatchR**, highlighting its full flexibility when applied programmatically in R. This example also focuses on Bull Trout, but extends beyond a single site to a broad spatial analysis across 226 locations in the upper Columbia River headwaters in Idaho.

We use the `idaho` dataset from Isaak, Luce, Chandler, Horan, & Wollrab (2018), which includes temperature data for these sites. To identify putative Bull Trout Spawning locations, we apply a filtering criterion based on mean August temperature, as outlined in Isaak, Young, Nagel, Horan, & Groce (2015), selecting only sites with mean August temperature </= 13 °C, a known thermal threshold for Bull Trout spawning suitability. The filtering process reduced the dataset to 139 sites potential spawning sites.

To predict hatch timing across these sites, we first set up the necessary models and data, using Bull Trout parameterization (for concision, we omit this set up here, but full details are available in the `paper.Rmd` file in the GitHub repository or in the Predict fish phenology: nested article on **hatchR**'s website) We then apply `predict_phenology()` across all 139 sites, running predictions for three representative spawn dates (Early: September 1st, Peak: September 15th, Late: September 30th).

By mapping predictions across this broad spatial extent, we generate a large-scale assessment of Bull Trout phenology, illustrating how hatch timing varies across different spawning habitats. The results of this analysis are presented in Figure 5, providing insights into how hatch timing might vary under different thermal regimes and across the species' geographic range. This case study underscores the power of **hatchR** for large-scale ecological applications, particularly in conservation planning and habitat suitability assessments.

```
hatch_res <- isaak_summ_bt |>
  mutate(
    dev_period = map2(
      summ_obj, spawn_dates, # map across our site object and spawn dates
      predict_phenology,
      temperature = daily_temp,
      model = bt_hatch,
      dates = date
      ) |>
      map_df("dev.period") |> # pull out just dev.period results
      list()
    ) |>
  select(site, dev_period) |> # just select the columns we want
  unnest(cols = c(dev_period)) |> # unnest everything
  mutate(days_to_hatch = stop - start) # make a new column of days to hatch
```

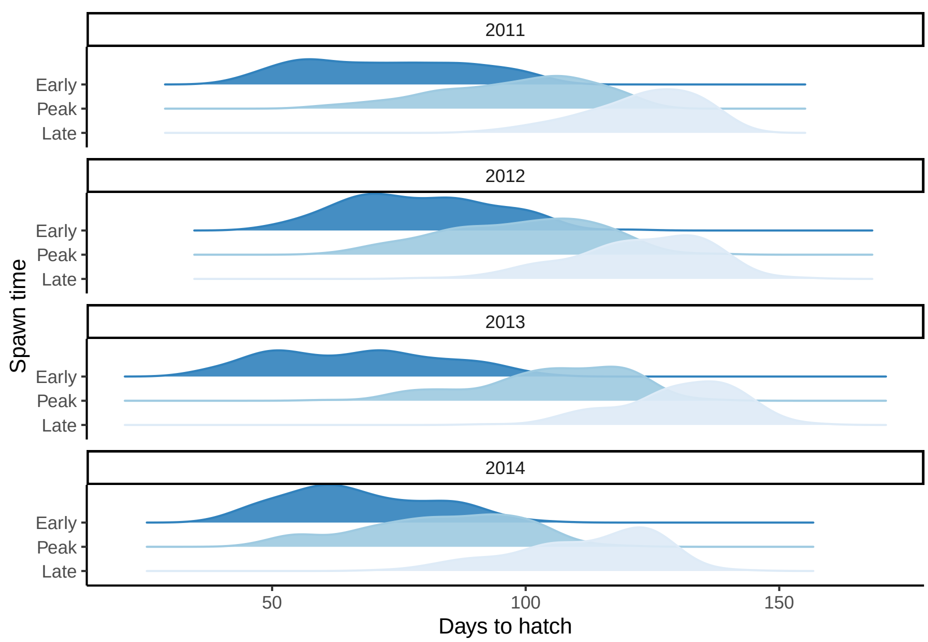

Figure 5: Predicted days to hatch for 139 putative bull trout populations over three spawning periods (Early = September 1, Peak = September 15, Late = September 30) and four years of temperature data.

Discussion

With **hatchR**, we present a software ecosystem that bridges the analytical gap in predicting developmental phenology for wild fishes. It establishes a formal framework for applying effective value models from user-provided parameterizations. The software is available in two formats: 1) A fully customizable R package, ideal for complex and repetitive analyses and 2) a graphical user interface for ease of use, designed for tasks that may only need to be run once or a few times.

Both versions allow users to import data, perform basic data checks, and apply either pre-existing salmonid model parameterizations or generate custom models specific to other species or populations. To support users at various levels of expertise, we provide extensive documentation on the hatchR website (https://bmait101.github.io/hatchR/), covering basic and advanced applications.

Assumptions and considerations in applying effective value models

The application of hatchR and the effective value modeling framework relies on several key assumptions. First, environmental stressors may alter developmental timing. While effective value models predict developmental timing based on temperature, studies have shown that stressful environmental condition such as low dissolved oxygen, altered pH, high salinity, pathogen exposure, or mechanical disturbance can induce premature hatching or emergence (Cowan et al., 2024; Quinn, 2018). Users should consider how such factors may influence their predictions.

Second, developmental timing occurs as a distribution, not a fixed point. While **hatchR** provides point estimates for developmental phenology, fish spawning and development within populations occur as distributions rather than single events (Mason, 1976). We encourage users to predict phenology using early, peak, and late thresholds (e.g., 5th, 50th, and 95th percentiles) or incorporate real or modeled distributions to capture variation.

Third, sensor-based temperature data may differ from actual embryonic ambient temperatures. Water temperatures recorded by environmental sensors may not fully reflect thermal conditions in spawning microhabitats, where geomorphic factors influence temperature regimes (Geist et al., 2002). Users should consider how differences between measured and actual incubation temperatures may affect predictions.

Evolutionary and population-level considerations

To date, effective value models have primarily been used to predict phenology in wild environments using species-specific parameterizations (Adelfio et al., 2024; Austin et al., 2019). However, these models fundamentally represent reaction norms, meaning that temperature-development relationships are influenced by local adaptation, gene-environment interactions, and phylogenetic differences (West-Eberhard, 2003).

For example, Sparks et al. (2017) found no significant differences in developmental rates between populations in their study but did observe family-level genetic × environment interactions under different thermal regimes. Similarly, when they reparameterized their models using western Alaskan Sockeye Salmon, they found slower developmental rates compared to populations from Canada (Beacham & Murray, 1990), consistent with cogradient variation (Conover, Duffy, & Hice, 2009; Sparks, Kraft, Blackstone, McNickle, & Christie, 2022). These findings highlight the importance of considering how developmental rates are keyed to specific environments but also how these underlying statistical relationships inform micro- and macro-evolutionary processes in fishes.

Expanding the utility of hatchR

The models within **hatchR** can be customized in multiple ways beyond the examples provided. While our current framework focuses on predicting hatch or emergence timing, it could be adapted to other key developmental milestones not reliant on exogenous feeding, such as early embryonic stages (*e.g.*, eye-up; (Velsen, 1980)), initiation or cessation of pelagic-larval dispersal, or current-mediated dispersal in riverine species, though not all cases may be as specifically tied to temperature as hatch and emergence.

Additionally, while `fit_model()` uses non-linear regression to estimate parameters, `predict_phenology()` only requires users to provide a model expression. This means that users can integrate alternative model structures, as long as they incorporate daily temperature, allowing further customization of predictions.

Finally, while hatchR was designed specifically for fishes, it has potential applications for other poikilothermic organisms, such as reptiles, amphibians, and invertebrates, where developmental rates similarly follow a power law relationship with temperature. Extending the effective value framework to these taxa could provide valuable insights into their life history timing under variable environmental conditions.

Conclusion

hatchR provides a versatile and accessible tool for predicting developmental phenology in wild fish populations. It offers basic data checks and summarization tools, pre-existing and customizable model parameterization options, and scalable applications from simple site-level predictions to complex multi-site analyses.

Importantly, hatchR extends the effective value framework developed by Sparks et al. (2019) into a generalizable tool that can be applied to any fish species or population, provided that appropriate source data are available—data that can easily be collected in aquaculture settings. We present foundational applications of **hatchR**, with additional use cases and implementation guides available on the software's website. The software is designed for both applied and basic research, allowing users to engage with it either through a programmatic R environment or via a user-friendly Shiny app. We expect that the examples provided here represent only a fraction of **hatchR**'s potential applications and encourage the user community to explore and expand upon this framework for their own research and management needs.

Acknowledgements

We thank Laura Koller for her help designing the **hatchR** logo. Dan Isaak provided helpful discussion about model development and temperature data sets.

The views expressed in this manuscript are those of the authors and do not necessarily represent the views or policies of USFS. Any mention of trade names, products, or services does not imply an endorsement by the U.S. government or USFS. USFS does not endorse any commercial products, services or enterprises.

Conflicts of Interest

The authors declare no known conflicts of interest.

Data Availability

hatchR is fully open source and reproducible. Source code and data can be found at https://github.com/bmait101/hatchR/. The Rmardkdown document with all the code to reproduce the examples from this manuscript is available at https://github.com/bmait101/hatchR/blob/master/inst/manuscript/paper.Rmd. The The latest version will be archived upon acceptance of the manuscript.

Ethics Statement

All data was derived from pre-published sources or created synthetically.

Funding

The authors declare no funding sources.

References

Adelfio, L. A., Wondzell, S. M., Mantua, N. J., & Reeves, G. H. (2019). Warm winters reduce landscape-scale variability in the duration of egg incubation for coho salmon (oncorhynchus kisutch) on the copper river delta, alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, *76*(8), 1362–1375. doi:10.1139/cjfas-2018-0152

Adelfio, L. A., Wondzell, S. M., Mantua, N. J., & Reeves, G. H. (2024). Expanded, compressed, or equal? Interactions between spawning window and stream thermal regime generate three responses in modeled juvenile emergence for pacific salmon. *Canadian Journal of Fisheries and Aquatic Sciences*. doi:10.1139/cjfas-2023-0238

Alderdice, D. F., & Velsen, F. P. J. (1978). Relation between temperature and incubation time for eggs of chinook salmon (oncorhynchus tshawytscha). *Journal of the Fisheries Research Board of Canada*, *35*(1), 69–75. doi:10.1139/f78-010

Austin, C. S., Essington, T. E., & Quinn, T. P. (2019). Spawning and emergence phenology of bull trout Salvelinus confluentus under differing thermal regimes. *Journal of Fish Biology*, *94*(1), 191–195. doi:10.1111/jfb.13864

Beacham, T. D., & Murray, C. B. (1990). Temperature, egg size, and development of embryos and alevins of five species of pacific salmon: A comparative analysis. *Transactions of the American Fisheries Society*, *119*(6), 927–945. doi:10.1577/1548-8659(1990)119<0927:TESADO>2.3.CO;2

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., et al. (2024). *Shiny: Web application framework for r*. Retrieved from https://cran.r-project.org/web/packages/shiny/index.html

Conover, D. O., Duffy, T. A., & Hice, L. A. (2009). The Covariance between Genetic and Environmental Influences across Ecological Gradients. *Annals of the New York Academy of Sciences*, *1168*(1), 100–129. doi:10.1111/j.1749-6632.2009.04575.x

Cowan, Z.-L., Green, L., Clark, T. D., Blewett, T. A., De Bonville, J., Gagnon, T., Hoots, E., et al. (2024). Global change and premature hatching of aquatic embryos. *Global Change Biology*, *30*(9), e17488. doi:10.1111/gcb.17488

Geist, D. R., Hanrahan, T. P., Arntzen, E. V., McMichael, G. A., Murray, C. J., & Chien, Y.-J. (2002). Physicochemical Characteristics of the Hyporheic Zone Affect Redd Site Selection by Chum Salmon and Fall Chinook Salmon in the Columbia River. *North American Journal of Fisheries Management*, *22*(4), 1077–1085. doi:10.1577/1548-8675(2002)022<1077:PCOTHZ>2.0.CO;2

Isaak, D. J., Luce, C. H., Chandler, G. L., Horan, D. L., & Wollrab, S. P. (2018). Principal components of thermal regimes in mountain river networks. *Hydrology and Earth System Sciences*, *22*(12), 6225–6240. doi:10.5194/hess-22-6225-2018

Isaak, D. J., Young, M. K., Nagel, D. E., Horan, D. L., & Groce, M. C. (2015). The cold-water climate shield: delineating refugia for preserving salmonid fishes through the 21st century. *Global Change Biology*, *21*(7), 2540–2553. doi:10.1111/gcb.12879

Kaylor, M. J., Justice, C., Armstrong, J. B., Staton, B. A., Burns, L. A., Sedell, E., & White, S. M. (2021). Temperature, emergence phenology and consumption drive seasonal shifts in fish growth and production across riverscapes. *Journal of Animal Ecology*, *90*(7), 1727–1741. doi:10.1111/1365-2656.13491

Mason, J. (1976). Some features of coho salmon, oncorhynchus kisutch, fry emerging from simulated redds and concurrent changes in photobehavior. *Fish. Bull*, *74*(1), 167–175.

McPhail, J., & Murray, C. (1979). *The early life-history and ecology of dolly varden (salvelimus malma) in the upper arrow lakes*. Nelson, BC, Canada.

Nolfi, D., Melbihess, T., Fisher, S., & Ellis, L. (2024). 5-Year Statuse Review Coterminous United States Population of Bull Trout (Salvelinus confluentus).

Quinn, T. P. (2018). *The Behavior and Ecology of Pacific Salmon and Trout*. University of Washington Press.

R: The r project for statistical computing. (n.d.). Retrieved from https://www.r-project.org/

Siegel, J. E., Fullerton, A. H., FitzGerald, A. M., Holzer, D., & Jordan, C. E. (2023). Daily stream temperature predictions for free-flowing streams in the Pacific Northwest, USA. *PLOS Water*, *2*(8), e0000119. doi:10.1371/journal.pwat.0000119

Small, B. C., & Bates, T. D. (2001). Effect of Low-Temperature Incubation of Channel Catfish Ictalurus punctatus Eggs on Development, Survival, and Growth. *Journal of the World Aquaculture Society*, *32*(2), 189–194. doi:10.1111/j.1749-7345.2001.tb01094.x

Smith, K. M., & King, D. K. (2005). Dynamics and extent of larval lake sturgeon Acipenser fulvescens drift in the Upper Black River, Michigan. *Journal of Applied Ichthyology*, *21*(3), 161–168. doi:10.1111/j.1439-0426.2005.00623.x

Sparks, M. M., Falke, J. A., Quinn, T. P., Adkison, M. D., Schindler, D. E., Bartz, K., Young, D., et al. (2019). Influences of spawning timing, water temperature, and climatic warming on early life history phenology in western alaska sockeye salmon. *Canadian Journal of Fisheries and Aquatic Sciences*, *76*(1), 123–135. doi:10.1139/cjfas-2017-0468

Sparks, M. M., Kraft, J. C., Blackstone, K. M. S., McNickle, G. G., & Christie, M. R. (2022). Large genetic divergence underpins cryptic local adaptation across ecological and evolutionary gradients. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1984), 20221472. doi:10.1098/rspb.2022.1472

Sparks, M. M., Westley, P. A. H., Falke, J. A., & Quinn, T. P. (2017). Thermal adaptation and phenotypic plasticity in a warming world: Insights from common garden experiments on Alaskan sockeye salmon. *Global Change Biology*, *23*(12), 5203–5217. doi:10.1111/gcb.13782

Velsen, F. P. J. (1980). Embryonic Development in Eggs of Sockeye Salmon Oncorhynchus nerka. *Canadian Special Publication of Fisheries and Aquatic Sciences*, (49).

Webster, D. A. (1948). Relation of temperature to survival and incubation of the eggs of smallmouth bass (micropterus dolomieu). *Transactions of the American Fisheries Society*, *75*(1), 43–47. doi:10.1577/1548-8659(1945)75[43:ROTTSA]2.0.CO;2

West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford University Press.