

20-March-2017

Below are step-wise descriptions of example data sets and Program R code for calculating secondary production for taxa with distinguishable cohort structure. This method is commonly referred to as the instantaneous growth method ($\text{mg dry mass m}^{-2} \text{ day}^{-1}$; Benke and Huryn 2006). A bootstrapping technique is used to estimate uncertainty around means (Benke and Huryn 2006).

- Email questions, comments, and ideas on how to improve either the analysis itself or the methodological description to Michael Venarsky (mvenarsky@gmail.com). Meaningful contributions will be credited to the person.

Data files

Length_mass_v1.csv description:

- TAXA = a unique number assigned to each taxa used to link data tables (see Cohort_v1.csv and Remove_date_v1.csv).
- 1 to 35 = Size classes (mm). These columns contain the biomass for one individual (mg dry mass).
- Length mass equations were acquired from the literature (Callow, 1975; Culver et al., 1985; Leeper & Taylor, 1998; Benke et al., 1999; Doroszuk et al., 2007; Lemke & Benke, 2009).

Remove_date_v1.csv description:

- TAXA = a unique number assigned to each taxa used to link data tables (see Cohort_v1.csv and Length_mass_v1.csv).
- GENUS = name assigned to each taxa used to link data tables (see Cohort_v1.csv and Length_mass_v1.csv).
- DATE = sampling date
- CROP = "1" in this column will remove the date from the instantaneous growth rate calculations.
 - Cohorts appear and disappear during a study because larva mature and adult insects emerge. Dates when cohorts were not present were removed so that calculations did not return negative instantaneous growth rate values, which can lead to secondary production rates.

Cohort_v1.csv description:

- DATE = sampling date
- HABITAT = type of habitat samples. Pool is depositional habitat dominated by fine substrates. Riff is erosional habitat dominated by large cobbles and boulders. Wood is sticks and logs submerged in the stream.

- TAXA = a unique number assigned to each taxa used to link data tables (see Length_mass_v1.csv and Remove_date_v1.csv).
- GENUS = name assigned to each taxa used to link data tables (see Length_mass_v1.csv and Remove_date_v1.csv).
- 1 to 23 = Size classes (mm). Note that the largest individual is 20mm. The excess size classes (21 to 23) will be removed during analyses. These columns contain the number of individuals per unit area (m²).
- Each line is a replicate. Note that replicates are NOT uniquely identified (e.g., Pool-1, Pool-2, etc.). Doing this will cause the code to return a “resampled data table” that is identical to the original data set.
- The tricopterans *Rhyacophila* and *Arctopsyche* have one (*Rhyacophila_2*) and two (*Arctopsyche_1* and *Arctopsyche_2*) distinguishable cohorts, respectively.
- Small values: 0.0001
 - Community data sets often have many zeros because every taxa from every cohort is not present in every sample. Thus, a resampling procedure could randomly select only samples that are zeros. This artifact of resampling can cause negative instantaneous growth rates, which leads to negative rates of secondary production. To avoid this issue, we inserted 0.0001 into the data sets to ensure that cohort structure is maintained during the resampling procedure.
 - These values are placed into the mean size class for a cohort.
 - These values do not significantly increase estimates of biomass or secondary production.

How we identified cohort structure:

- Summed all individuals per date and then used "Sparklines" columns to visualize cohorts (Excel: Insert tab, Sparklines, Column). Calculated weighted average to see how average body size increased over time. See excel table below.

DATE	TAXA	GENUS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Weighted average	Sparklines	
30-Jun-13	43	Arctopsyche_1	7	34																			1.83		
27-Jul-13	43	Arctopsyche_1																							
6-Nov-13	43	Arctopsyche_1				40	40	40	20	20													5.63		
30-Jun-13	43	Arctopsyche_2								7	20	20			20								10.39		
27-Jul-13	43	Arctopsyche_2													20		20	20	20	20		19	16.47		
6-Nov-13	43	Arctopsyche_2																							
30-Jun-13	51	Rhyacophila_2		40	80	200	107																3.88		
27-Jul-13	51	Rhyacophila_2			60	160		49	40														4.51		
6-Nov-13	51	Rhyacophila_2				40	100	64	20														5.29		

Calculating secondary production

2oProd_cohort_v1.R:

Pre-analysis setup:

- Code checked 20-March-2017
- Constructed using R version 3.3.1 (2016-06-21) -- "Bug in Your Hair". Copyright (C) 2016 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit).
- This code is based on an excel macro constructed by Prof. Alex Huryn (University of Alabama, Department of Biological Sciences).
- Cam MacKenzie originally wrote code for Steps 1 to 3 (taxa with distinguishable cohorts) and Steps 2 to 3 (taxa without distinguishable cohorts) using Program R base package. Michael Venarsky rewrote these steps using dplyr, tidyr, lubridate, and zoo packages on 28-March-2015. Several stackoverflow.com users provided assistance to Michael Venarsky. See questions by stackoverflow user Vesuccio.
- For those not familiar with dplyr or tidyr, consider loading the swirl package which is an interactive tutorial for dplyr and tidyr implemented in R studio.
- Links to dplyr and tidyr cheat sheets:
 - <http://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf>
 - <https://www.rstudio.com/resources/cheatsheets/>
- Include three folders your working directory: data, results, check. CSV files will be written to each of these folders. Note that CSV files are written to the "check" folder only when you delete the "#" at the beginning of the line of code.
- Clear all data from previous runs, load required packages, set working directory and import all data files.

Step 1

- Resample_function resamples Abundance with replacement.

Step 2

- N_resample defines the number of resampling events. We conducted 1000.
- Beginning of Loop 1, which is a very long loop. Loop 1 begins by running the Resample_function function.
- Size_class and Loop 2 remove size classes with 0 individuals from Resample_abundance.

Step 3

- Daily instantaneous growth rate calculations. Calculations in this step are NON-HABITAT-SPECIFIC because many taxa are habitat-specific. Variations in the density and size of individuals among habitats is likely due to habitat selection rather than higher or lower growth rates associated with specific habitats.
- Step 3.1 - Calculates average abundance of individuals per size class per taxa per date.
- Step 3.2 - Converts abundance to biomass (mg dry mass) by joining Avg_abundance_size and Length_mass tables. Calculates the average biomass of one individual per taxa per date and then natural log transforms biomass. Calculates number of days between sampling dates (DAYSB) and total number of days elapsed since first sampling date (DAYS_ELAPSED).
- Step 3.3 – Removes sampling dates from growth rate calculations by joining Remove_date and Avg_biomass_size tables. Calculates daily instantaneous growth rate (IGR) and the regression intercept (INTERCEPT) by regressing natural log of biomass with number of days elapsed.

Step 4

- Calculations in this step are NON-HABITAT-SPECIFIC (see Step 3 for explanation).
- Step 4.1 – Joins Avg_biomass_size and Growth_rate tables and then calculates interval secondary production (PINT) by averaging biomass between sampling dates then multiplying by the instantaneous daily growth rate (IGR) and the number of days between sampling dates (DAYSB). Units are mg dry mass m⁻² d⁻¹
- Step 4.2 - Calculates total secondary production (P) and cohort secondary production to biomass ration (P/B) by summing interval secondary production (PINT) and then dividing by mean biomass (B).
- *Daily instantaneous growth rate (IGR) and cohort secondary production to biomass ration (P/B) were extracted and used to calculate secondary production for taxa with indistinguishable cohorts (see 2oProd_IGR_no_cohort_v1 and 2oProd_PB_no_cohort_v1).*

Step 5

- Habitat-specific secondary production. Calculations in this step are HABITAT-SPECIFIC.
- Step 5.1 - - Calculates average abundance of individuals per size class per taxa per date.
- Step 5.2 - Converts abundance to biomass (mg dry mass) by joining Hab_avg_abundance_size and Length_mass tables. Calculates average biomass per taxa per habitat per date.

- Step 5.3 – Joins Hab_avg_biomass_size and Growth_rate tables and then calculates interval secondary production (PINT) by averaging biomass between sampling dates then multiplying by the instantaneous daily growth rate (IGR) and the number of days between sampling dates (DAYSB). Units are mg dry mass m⁻² d⁻¹.
- Step 5.4 – Calculates total secondary production (P) and mean biomass (B) per HABITAT and then joins cohort secondary production to biomass ratio (P/B), daily instantaneous growth rate (IGR), and the regression intercept (INTERCEPT) from Step 4.2.

Step 6

- N_resample defines the number of times Loop 1 is run. We conducted 1000. If the N_resample is 1, the output from the analyses are exported. If the N_resample is >1, then the results from each resampling are sequentially assembled into a table and then exported as a CSV file.
- Both NON-HABITAT-SPECIFIC and HABITAT-SPECIFIC results are exported as CSV files.

Step 7

- Weak cohort structure can cause some resampling events to produce negative instantaneous growth rates and thereby negative secondary production values. There are two ways to deal with this issue: A) Restart the resampling run when negative growth rates are generated; B) After 1000 resampling events, remove the negative secondary production values and then resample the remaining secondary production values to replace the deleted values. This will preserve the mean of non-negative secondary production values. We chose to execute option B, but we believe either option is appropriate. When more than 5% of the resampled data sets produce negative secondary production values, the taxa were considered to have weak cohort structure and were placed into “without distinguishable cohorts” data set (see below).
- Step 7.1 - Removes negative secondary production (P) values from Site_prod and Hab_prod
- Step 7.2 - Resamples the positive secondary production values to bring the total bootstrap data set back to 1000. RESAMPLE column indicates the number of resampling events.

Step 8

- Calculate 95% quantiles for both NON-HABITAT-SPECIFIC and HABITAT-SPECIFIC data.

Literature cited

Benke, A. and A. Huryh (2006). "Secondary production of macroinvertebrates" *in* Methods in stream ecology. Hauer FR, Lamberti GA editors. Pages 691-710.

Benke, A. C., et al. (1999). "Length-mass relationships for freshwater macroinvertebrates in North America with particular reference to the southeastern United States." *Journal of the North American Benthological Society*: 308-343.

Calow, P. (1975). "Length-dry weight relationships in snails: some explanatory models." *Journal of Molluscan Studies* 41(4): 357-375.

Culver, D. A., et al. (1985). "Biomass of freshwater crustacean zooplankton from length-weight regressions." *Canadian Journal of Fisheries and Aquatic Sciences* 42(8): 1380-1390.

Doroszuk, A., et al. (2007). "Response of secondary production and its components to multiple stressors in nematode field populations." *Journal of Applied Ecology* 44(2): 446-455.

Leeper, D. A. and B. E. Taylor (1998). "Abundance, biomass and production of aquatic invertebrates in Rainbow Bay, a temporary wetland in South Carolina, USA." *Archiv für Hydrobiologie* 143(3): 335-362.

Lemke, A. and A. C. Benke (2009). "Spatial and temporal patterns of microcrustacean assemblage structure and secondary production in a wetland ecosystem." *Freshwater Biology* 54(7): 1406-1426.