Comparison of stripped down BASEmetab with original

Jan. 4th, 2021

This document provides a brief comparison of metabolism results obtained from the extracted BASEmetab code (“stripped down” hereafter) and those from the original [BASEmetab package](https://github.com/dgiling/BASEmetab). Data and R code for the stripped down version are available [here](https://github.com/fawda123/BASEmetab_script).

## Background

The original BASEmetab package includes a single function bayesmetab() that processes input data files for use with the JAGS model (i.e., [BASE\_metab\_model\_v2.3.txt](https://github.com/dgiling/BASEmetab/blob/master/inst/BASE_metab_model_v2.3.txt)). This function includes routines for data pre-processing, QA/QC checks for model results, and extraneous output not relevant to our questions. We attempted to extract the relevant code from the bayesmetab() function to simplify the core components used by the JAGS model and to streamline processing for use with multiple datasets. This simplified version is available as a standalone R script [here](https://github.com/fawda123/BASEmetab_script/blob/master/script.R) and a simplified JAGS model [here](https://github.com/fawda123/BASEmetab_script/blob/master/BASE_metab_model_v2.3.txt).

Modifications to the standalone R script include:

* Extraction of relevant arguments from the core bayesmetab() function that are only necessary to execute the stripped down code.
* Simplification of output results for key metabolism components, including gas exchange, GPP, ER, and NEP.
* Individual days are processed using multiple cores to reduce computation time.

Modifications to the JAGS model include:

* DO saturation is estimated at each time step prior to running the model (previously this was done as a separate loop inside the JAGS model). DO saturation is estimated using the observed data in the standalone R script with a helper function available [here](https://github.com/fawda123/BASEmetab_script/blob/master/R/dosat_fun.R).
* An additional loop that calculated instantaneous metabolic estimates was combined with the main loop for estimating dissolved oxygen mass balance.

An option to update.chains is also available in the bayesmetab() function as part of the JAGS convergence procedure that uses additional Monte Carlo sampling if results are unstable. This option was not used (i.e., update.chains = F) when comparing the stripped down version with BASEmetab because it was not retained in the former. It should be included in future analyses to ensure the results are robust.

Finally, the original BASEmetab method estimates gas exchange using prior distributions in a Bayesian framework. Both the stripped down version and the BASEmetab results here do not estimate gas exchange with Bayesian statistics to simplify our anticipated comparison with the Odum open-water method. Instead, both approaches use an empirically estimated gas exchange value following the Wanninkhof method. Currently, the same gas exchange estimate is applied to all days.

## Comparison of results

A four day [Appalachicola test dataset](https://github.com/fawda123/BASEmetab_script/blob/master/output/APNERR2012dtd.csv) was used to compare results between the stripped down version and the original BASEmetab output. This sample dataset uses “detided” dissolved oxygen to estimate metabolism. A full comparison of the methods and results can be done by [cloning](https://docs.github.com/en/free-pro-team@latest/github/creating-cloning-and-archiving-repositories/cloning-a-repository) the [GitHub repository](https://github.com/fawda123/BASEmetab_script) that includes all functions and datasets in a portable directory, then running the metbolism functions on a local computer. Once the repository is available locally, the following steps can reproduce the comparison of results below.

1. Run the code in [script.R](https://github.com/fawda123/BASEmetab_script/blob/master/script.R) to obtain results from stripped down version
2. Run the code in [script2.R](https://github.com/fawda123/BASEmetab_script/blob/master/script2.R) to obtain results from the original BASEmetab package
3. Run the examples below to view the comparison

The results from the stripped down version are available in the repository as a .RData object file called [outputpar](https://github.com/fawda123/BASEmetab_script/blob/master/data/outputpar.RData). The results from the BASEmetab original method are available as a .csv file called [BASe\_results\_2020-12-21 161657.csv](https://github.com/fawda123/BASEmetab_script/blob/master/output/BASE_results_2020-12-21%20161657.csv). For simplicity, they are named strpdat and origdat for the stripped down and original results in this document. Note that origdat includes many more columns than strpdat as part of the original output from BASEmetab.

strpdat

## Date GPP ER NEP K  
## 1 5/1/2012 0.1177762 0.3053659 -0.1875897 0.4313938  
## 2 5/2/2012 0.2634145 0.4358824 -0.1724679 0.4313938  
## 3 5/3/2012 0.4526297 0.6836766 -0.2310469 0.4313938  
## 4 5/4/2012 0.4828993 0.7564247 -0.2735253 0.4313938

origdat

## X File Date GPP.mean GPP.sd GPP.median ER.mean  
## 1 1 APNERR2012dtd.csv 5-1-2012 0.1177762 0.04254762 0.1158657 0.3053659  
## 2 2 APNERR2012dtd.csv 5-2-2012 0.2634145 0.03552543 0.2636721 0.4358824  
## 3 3 APNERR2012dtd.csv 5-3-2012 0.4526297 0.02132614 0.4528560 0.6836766  
## 4 4 APNERR2012dtd.csv 5-4-2012 0.4828993 0.02500584 0.4828848 0.7564247  
## ER.sd ER.median NEP.mean NEP.sd NEP.median PR.mean PR.sd  
## 1 0.04898103 0.3037444 -0.1875897 0.015613245 -0.1874168 0.3736119 0.08948717  
## 2 0.04066986 0.4360910 -0.1724679 0.012731888 -0.1724335 0.6022904 0.03331625  
## 3 0.02533534 0.6836775 -0.2310469 0.007957271 -0.2310598 0.6618556 0.01057452  
## 4 0.02814585 0.7563789 -0.2735253 0.009367782 -0.2734529 0.6381184 0.01320429  
## PR.median K.mean K.sd K.median theta.mean theta.sd theta.median  
## 1 0.3831762 0.4313938 9.997215e-10 0.4313938 1.07177 0 1.07177  
## 2 0.6037264 0.4313938 1.005497e-09 0.4313938 1.07177 0 1.07177  
## 3 0.6617708 0.4313938 1.005497e-09 0.4313938 1.07177 0 1.07177  
## 4 0.6381336 0.4313938 1.005497e-09 0.4313938 1.07177 0 1.07177  
## A.mean A.sd A.median p.mean p.sd p.median R2  
## 1 1.892902e-06 6.838267e-07 1.862197e-06 1 0 1 0.2822506  
## 2 4.651528e-06 6.273289e-07 4.656078e-06 1 0 1 0.5737844  
## 3 1.182442e-05 5.571204e-07 1.183033e-05 1 0 1 0.8792681  
## 4 9.641895e-06 4.992836e-07 9.641606e-06 1 0 1 0.8287021  
## PPP rmse rmse.relative mrl.fraction ER.K.cor convergence.check  
## 1 0.3693333 0.07090204 9.815392 0.5520833 0.01613546 Fine  
## 2 0.3703333 0.05746505 8.188587 0.5104167 -0.01559251 Fine  
## 3 0.3796667 0.03421807 4.708832 0.2916667 -0.01648518 Fine  
## 4 0.4400000 0.04258639 5.779079 0.3437500 -0.01371530 Fine  
## A.Rhat K.Rhat theta.Rhat p.Rhat R.Rhat GPP.Rhat DIC pD  
## 1 1.022706 1 1 1 1.002226 1.022706 -226.3608 2.925454  
## 2 1.000724 1 1 1 1.001133 1.000724 -266.0093 3.106597  
## 3 1.000739 1 1 1 1.001240 1.000739 -364.2833 3.310835  
## 4 1.000756 1 1 1 1.001259 1.000756 -322.7098 3.341292  
## totDailyLight aveDailyTemp interval smooth.DO smooth.PAR n.iter n.burnin  
## 1 62219.89 25.84062 900 0 FALSE 20000 10000  
## 2 56629.67 26.12083 900 0 FALSE 20000 10000  
## 3 38279.22 26.12292 900 0 FALSE 20000 10000  
## 4 50083.44 26.19583 900 0 FALSE 20000 10000

The core elements of origdat to compare with strpdat include GPP, ER, NEP, and K. These are selected and renamed from origdat to simplify the output for comparison to strpdat.

library(dplyr)  
origdat <- select(origdat, Date, GPP = GPP.mean, ER = ER.mean, NEP = NEP.mean, K = K.mean)  
origdat

## Date GPP ER NEP K  
## 1 5-1-2012 0.1177762 0.3053659 -0.1875897 0.4313938  
## 2 5-2-2012 0.2634145 0.4358824 -0.1724679 0.4313938  
## 3 5-3-2012 0.4526297 0.6836766 -0.2310469 0.4313938  
## 4 5-4-2012 0.4828993 0.7564247 -0.2735253 0.4313938

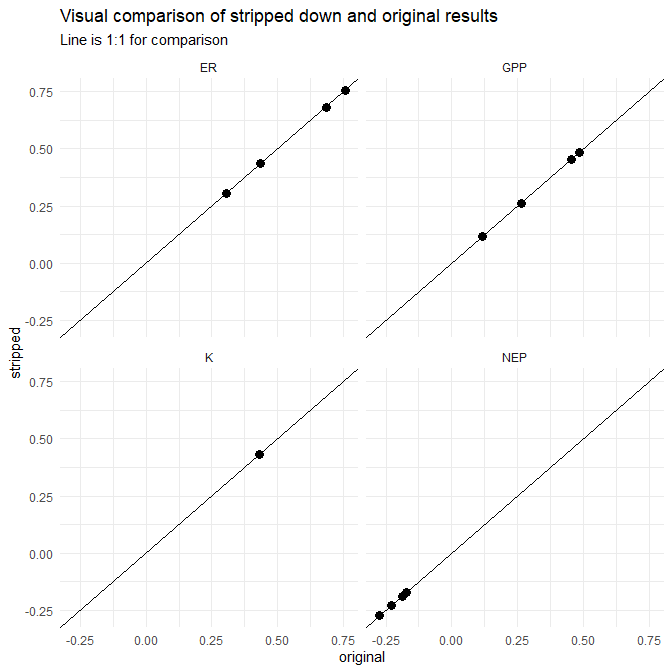
The two data objects can also be plotted to verify identical results, but first some data wrangling.

library(ggplot2)  
library(tidyr)  
library(lubridate)  
  
strpdat <- strpdat %>%   
 mutate(Date = mdy(Date)) %>%   
 gather('var', 'stripped', -Date)  
origdat <- origdat %>%   
 mutate(Date = mdy(Date)) %>%   
 gather('var', 'original', -Date)  
toplo <- full\_join(strpdat, origdat, by = c('Date', 'var'))  
toplo

## Date var stripped original  
## 1 2012-05-01 GPP 0.1177762 0.1177762  
## 2 2012-05-02 GPP 0.2634145 0.2634145  
## 3 2012-05-03 GPP 0.4526297 0.4526297  
## 4 2012-05-04 GPP 0.4828993 0.4828993  
## 5 2012-05-01 ER 0.3053659 0.3053659  
## 6 2012-05-02 ER 0.4358824 0.4358824  
## 7 2012-05-03 ER 0.6836766 0.6836766  
## 8 2012-05-04 ER 0.7564247 0.7564247  
## 9 2012-05-01 NEP -0.1875897 -0.1875897  
## 10 2012-05-02 NEP -0.1724679 -0.1724679  
## 11 2012-05-03 NEP -0.2310469 -0.2310469  
## 12 2012-05-04 NEP -0.2735253 -0.2735253  
## 13 2012-05-01 K 0.4313938 0.4313938  
## 14 2012-05-02 K 0.4313938 0.4313938  
## 15 2012-05-03 K 0.4313938 0.4313938  
## 16 2012-05-04 K 0.4313938 0.4313938

Now a simple plot to verify the comparison.

ggplot(toplo, aes(x = original, y = stripped)) +   
 geom\_abline(intercept = 0, slope = 1) +  
 geom\_point(size = 3) +   
 facet\_wrap(~var, ncol = 2) +  
 theme\_minimal() +   
 labs(  
 title = 'Visual comparison of stripped down and original results',  
 subtitle = 'Line is 1:1 for comparison'  
 )



## Next steps

This document is a simple comparison of only four days of data to verify successful extraction of the core components of BASEmetab. Following this comparison, additional analyses will include:

1. Comparing results for the entire time series of relevant data (i.e., not just four example days)
2. Including an option for update.chains in the stripped down version and comparing with the original BASEmetab results
3. Using a “metabolic” day to estimate metabolism from sunrise to sunrise the following day
4. Using instantaneous gas exchange estimates at each time step as compared to one estimate for the entire time series
5. Bayesian estimation of the contribution of tidal advection on the dissolved oxygen time series