CMEE miniproject

Model selection on the abundance of freshwater microbial under the condition of global warming

Yuxin Qin

Mar 2019

Department of life sciences (Silwood Park)

Imperial College London

yq3018@imperial.ac.uk

Word Count: 2713

Abstract

The increasing temperature of atmosphere has threaten the freshwater microbial communities due to the vulnerability, sensitivity and dependancy to temperatue of the freshwater systems. Reduced body-size is an obvious biological response to the warming condition in many ectothermic taxa for warming significant affects the life history and metabolic mechanism. Yvon and Stewart conducted similar meso-cosm experiments to test the impact of warming condition on freshwater microbial communities, while they came to different conclusions. Yvon claimed that community biomass declined consideratly under warming condition showing that the slope of log(abundance) against log(bodysize) in warming condition was steeper. However, Stewart's experiment came to an inconsistent conclusion which cannot prove the statement of Yvon. The aim of this article was to carry out the model selection by reanalyzing Stewart's data and comparing the allometric scaling model and to Plante-Downing model. It turned out that the Plante-Downing model performed better than the allometric scaling model in Akaike information criterion (AIC), while the allometric scaling model provided us valuable information on understanding the impacts of global warming on freshwater microbial communities.

Introduction

According to the IPCC report published in 2018, the surface temperature of the earth has increased approximately $1.0\,°C$ compared to pre-industrial levels and it tends to reach $1.5\,°C$ between 2030 and 2052 if global warming continues at current rate in the following years [6]. Since the phrase global warming first appeared in 1957, many scientists have studied the impacts of global warming in different scales and different organisms such as climate change, crop production, coral bleach and so on [15]. Due to temperature dependency and the long-term exposure to environmental stressors and the vulnerability of the freshwater system, it is significant to study the impacts of global warming on it [2]. The previous study of global warming on freshwater system mainly foceused on the distribution of freshwater and the potential changes of freshwater fishes [3], while many recent studies discovered that warming has obvious impacts on freshwater microbial communities by means of mesocosm experiments.

Utilizing mesocosm experiments to examine the impacts of warming on freshwater microbial communities allows researchers to control temperature, a key factor, between the experimental group and the control group. Mesocosm experiments also provide freshwater microbial communities less differences to the real natural environment compared to other ordinary laboratory experiments. By means of mesocosm ex-

peiments, Yvon discovered that warming significantly reduced the total biomass of freshwater microbial community and increased the gradient steepness of community size spectrum [16]. Rebeca Stewart from the University of London carried out a similar mesocosm experiment as Yvon did during her PhD, whereas with a different conclusion. She claimed a contrary opinion that warming has not shown a significant impact on freshwater microbial community biomass and increased gradient steepness of the community size spectrum [12]. In order to prove the actual influences of warming on freshwater microbial communities, a reanalysis of Stewart's data has been carried out in this study.

Stewart's mesocosm experiment consisted of 2 groups of ponds, the control group and the experiment group. Each group contained 10 artificial outdoor ponds with the capacity of approximately $1m^3$ water. The water temperature in experiment ponds were being maintained between 3-5 $1.0\,°C$ by electronic heating elements. All the experiment ponds and the control ponds were arranged randomly by the block design. The biota seeded in the mesocosms were from the river Frome, Dorset, UK, containing 5 main groups of microbes, ciliate, algae, meiofauna, flagellate and heterothrophic protists. These microbes were seeded to the ponds 10 months prior to experiments for required natural colonisation. The sample of microbial communities was performed monthly from February 2009 to January 2010. 10 ml water samples collected from each pond using a sterile syringe at three depths of the pond, the water surface, the middle water column and the sediment surface respectively. Afterwards, 1 ml subsample was removed from each 10 ml water sample to a 1 ml Sedgwich-Rafter counting cell for further analysis using light microscopy. Stewart identified the group, family, genus, species and feeding type of each microbial subsample and recorded the shapes lively under microscopy. The length and width of each microbial individual containing in the 1 ml subsample have been measured by image analysis software ImageJ and Q capture. Therefore, the biovolume, abundance and biomass of each species or groups were generated from the original data.

The objective of this article is to analyze the relationship among abundance, bodymass and biomass in freshwater microbial communities under warming condition. Shrinking bodysize has been observed as a severe ecological response to climate change in a wide range of ectothermic taxa [14]. Bodysize is regarded as a fundamental ecological feature of many biological characteristics such as reproduction, population density and growth rate, which are in an allometric scaliing relationship with bodysize [8]. The allometric scaling of bodysize and biological characteristic are a general criteria applied in many systems including freshwater system [16]. Another model, Plante-Downing model, is an empirical model with the merit of prediction for the invertebrate population in freshwater system[9]. This model selection research intended to do the data fitting of freshwater microbial collected by Stewart, with the coincidence evaluation of the

two above models by means of Akaike information criterion (AIC).

Methods

1 Data handling

The data selected in this article are based on the thesis of Rebecca Stewart. Some suspectable mistakes have been observed in the original datasheet, and for the sake of well-suited candidate model to the analysis, filtered blank space in temperature have been refilled, abnormal data have been deleted, the biovolume, bodymass and biomass have been recalculated in each data recorded. Biovolume of each freshwater microbial was calculated in length and width measured by the microscope according to the shape recorded, and equations applied to calculate the biovolume were followed the paper of Hillerbrand's paper[4]. Bodymass referring to the individual mass, calculated by the equation $Bodymass = Biovolume * 0.14 * 1.1$ according to Julia Resis's paper [10]. Biomass referring to the carbon dry mass of freshwater taxa, was calculated by the mutiplication of 0.25 bodymass and abundance. Therefore, these convenient recognized data for the fitting models have been obtained.

2 Fitting model

Two models, the allometric scaling model and the Plante-Downing model, have been introduced in this study to fit the abundance and bodymass values. The result coincidence of these two models has been compared on accounts of Akaike information criterion (AIC).

2.1 Allometric scaling model

The allometric scaling model is shown below, where y is the characteristic of interest, x is bodysize. a and b here are constants. The constant a can be considered as the ratio of specific growth rate between y and x [5]. For the sake of better fitting, the abundance was assumed as the biological characteristic y and the bodymass as the individual bodysize x. The logarithm operation applied on both bodymass and abundance, afterwards, the gradient of this linear model indicates the growth rate of bodymass and abundance. The slope of this linear model can indicate the growth rate of bodymass and abundance.

$$y = bx^a \tag{1}$$

4

2.2 Plante-Downing model

The Plante-Downing model is an empirical model shown below, where P is the population of freshwater invertebrate, $\overline{B}$ is the annual mean population biomass, W_m is the maximum individual bodymass and T is the ambient surface temperature [9]. This model has been proved to fit freshwater invetebrate communities annually quite well. [9]. In the circumstance of this study, the model has been fitted in monthly span and replaced P in the original formula with abundance. The rest parameters were performed as the original ones.

$$\log(P) = a + b \log(\overline{B}) - c \log(W_m) + dT \tag{2}$$

2.3 Model selection

In terms of model selection methods used in publications, AIC and BIC (Bayesian information criterion) are the most popular methods showing that AIC accounts for 84% of all and BIC accounts for 14% [1]. AIC utilized the Kullback-Leibler information lost to measure the discrepancy between the true model and the approximating model [13]. BIC also called the Schwarz or SIC criterion. The major difference between AIC and BIC is that AIC can better simulate situations with a true model, but BIC fits the situations maninly based on a true model [1].

The AIC has been chosen as the criterion for comparison of these two models since the relationship of abundance and bodymass in freshwater microbial did not reveal by a true model but only one model using bodymass to predict the abundance of freshwater. The calculation of AIC was based on the Johnson's papaer [7]. The normal AIC applied to this study because neither the number of the free parameters p in two selected models exceeded $n/40$ (n is the sample size). The sample size is 4271, $n/40 = 4271/40 = 106.67$, and the number of free parameters in the allometric model is 2 and the number of free parameters in Plante-Downing model is 4, thus obviously, both 2 and 4 are smaller than the 106.67. Hence, the second order deriative of AIC, the AIC_c, was not necessarily to apply in this paper since it is normally used to correct small sample size.

The formula to calculate AIC is shown below. In this formula, $L(\hat{\theta}|y)$ is the likelihood of the model parameters given by the data. For models fitted by least squares with the usual assumptions, $L(\hat{\theta}|y) = -n/2\ln(RSS/n)$. RSS here is the residual sum of squares for a linear model and n is the sample size. p is the free parameter of the tested model.

$$AIC = -2\ln[L(\hat{\theta}|y)] + 2p = -2\ln[-n/2\ln(RSS/n)] + 2p \tag{3}$$

5

3 Computing languages

The major computing languages programmed in the miniproject are bash, Python and R. Bash was used to running the workflow as creating the report in LaTeX,running Python scripts and R scripts. The calculation, model fitting and plotting are all performed in R version 3.2.3 (2015-12-10). The model fittings in this article have been done by lm in R, and the plots were produced by ggplot2 basically. Functions have been created to generate the coefficient table and AIC table automatically in R. Thress reasons for selected R to do calculation, model fiting and plotting: at first, the size of raw data only occupied 7.4Mb which is proper for R; the calculation and model fitting in this miniproject are simple and are able to be handled by R efficiently; at last, R provides a better visualization than Python. As for the ANOVA analysis in the article, python 3.7 has been performed on account of the lack of P value in ANOVA analysis when using lmer to do ANOVA in R is more complex than using statsmodels in Python.

Results

1 Community biomass

According to Figure 1, it is obvious that the mean community biomass under warming condition is smaller than that of the cold or ambient condition. The ANOVA analysis shown in Table 1 has shown that the treatment and month both have considerable effects on community biomass. It can be concluded that the biomass of freshwater microbial community largely reduced under warming conditions.
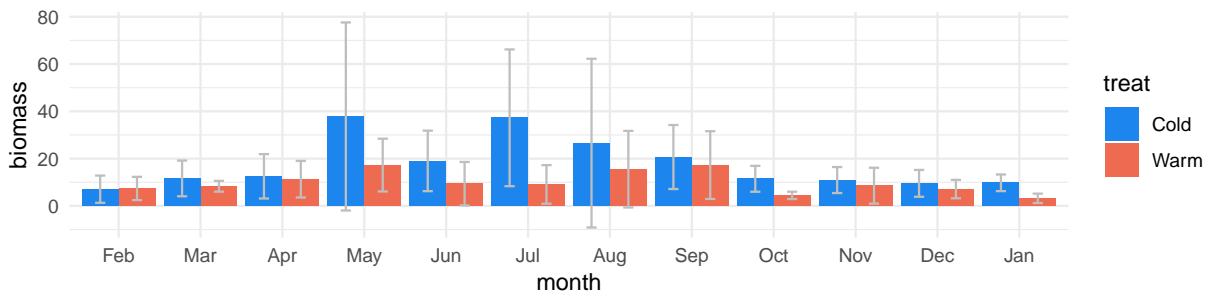

Figure 1: **The mean community biomass in each month.**

X	sum_sq	df	F	PR..F.
C(month)	11031.523	11	4.597698	2.784882e−06
C(treat)	3793.990	1	17.393774	4.400135e−05
C(month):C(treat)	3999.057	11	1.666719	8.252389e−02
Residual	47114.669	216	NA	NA

Table 1: **The ANOVA analysis of community biomass**

2 Allometric scaling model

The coefficients of the allometric scaling model under different conditions were listed in Table 2. By plotting the slope of warming condition and ambient condition on the same graph as shown in Figure 2, it is predictable that the slope of the ambient condition is higher than the warming condition except for February, the first experiment month. The slope of the linear model is also the coefficient of $log(bodymass)$, indicating a higher growth rate in population in the ambient condition compared to warming condition, which supports the conclusion driven by Yvon. The bodysize of ectothermic taxa mainly determined by two ecological factors, nutrient supply and metabolic rate [11]. Since the nutrient supply is the same among all mesocosm ponds, the reduction of bodymass in warming condition is attributed to the decreasing metabolic rate. Figure 2 implicated that one potential consequence of rising temperature is the decline of the metabolic rate of freshwater microbial communities.

By applying linear regression on the allometric scaling model show in Figure 3 and comparison of observed value against the predicted value of $log(abundance)$ shown in Figure 4, it is apparent the data did not express good fit into the model.

	Intercept	log.bodymass		Intercept	log.bodymass
Feb	1.06123173773668	−0.21736282407769	Feb	0.90650196087045	−0.265912509419171
Mar	1.11152305171191	−0.167997449554976	Mar	1.24646974044421	−0.119746539901104
Apr	1.05181018018766	−0.199653613125924	Apr	1.38335279929881	−0.134236755621018
May	1.21522448654769	−0.224576926748271	May	2.09411979589108	−0.0793231807800302
Jun	1.02682604741349	−0.218551436967741	Jun	1.8464434833877	−0.121713420203422
Jul	1.14382360225608	−0.195702290700023	Jul	2.1282970043348	−0.0801383978882819
Aug	1.29551126757694	−0.177691933152912	Aug	1.87149812807327	−0.0799498708767527
Sep	0.730992507117725	−0.253070444576018	Sep	1.49589333717046	−0.166439834630803
Oct	0.861726340911713	−0.192243067655338	Oct	1.46998227534853	−0.15064359655653
Nov	1.47126109002166	−0.109547425025442	Nov	1.60755904727647	−0.0066202965506389
Dec	1.3525783562076	−0.0588163509286233	Dec	1.65380934135525	0.025319660168804
Jan	0.936432271710948	−0.177571992140429	Jan	1.19140748418139	−0.123067457170078

Table 2: **The coefficients of the allometric scaling model.** The table on the left is the coefficients under warming condition. The table on the right is the coefficients under ambient condition.
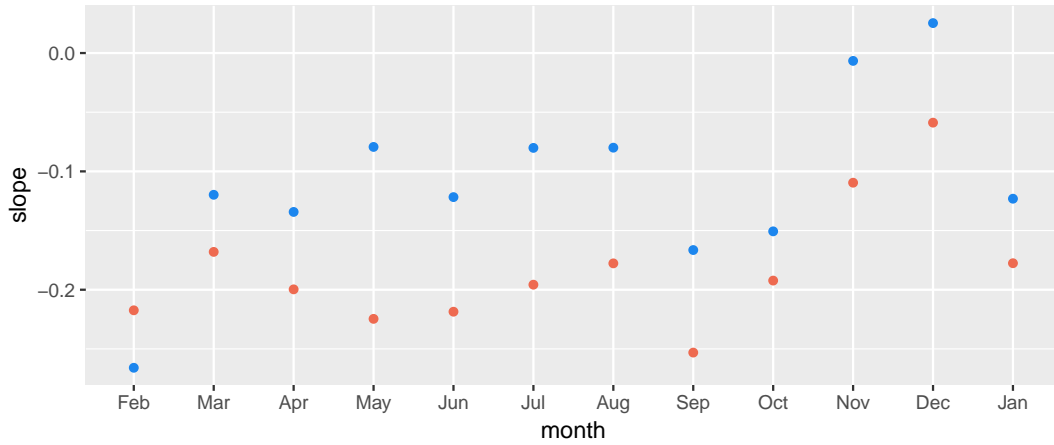

Figure 2: **The slope of allometric scaling model in each month.** The red dots represents samples of warming condition and the blue dots represents samples of ambient condition.
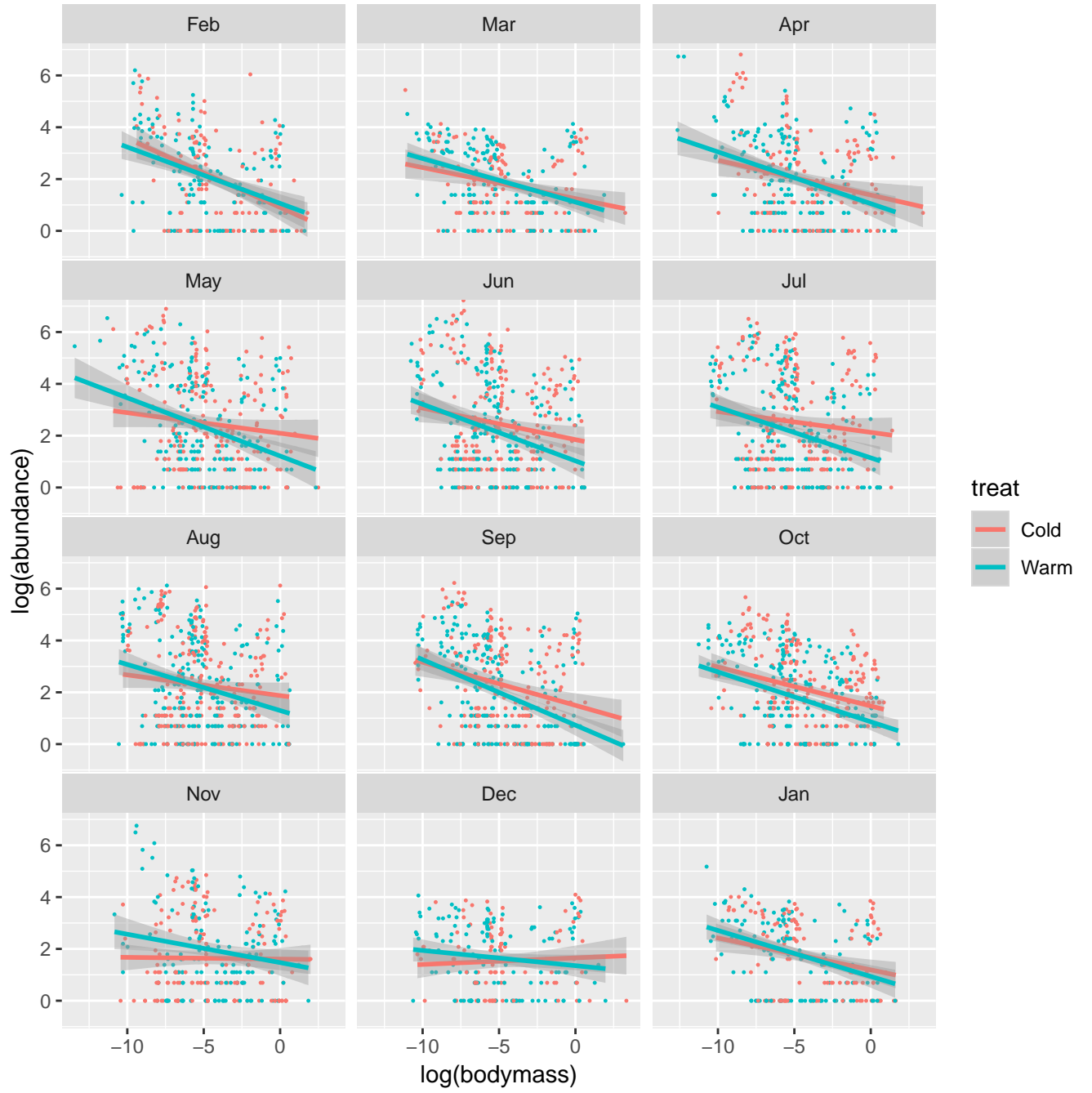
Figure 3: **Linear regression graph of the allometric scaling model.** Red scatters in the graph represent samples of warming condition and the blue scatters represent samples of ambient condition.
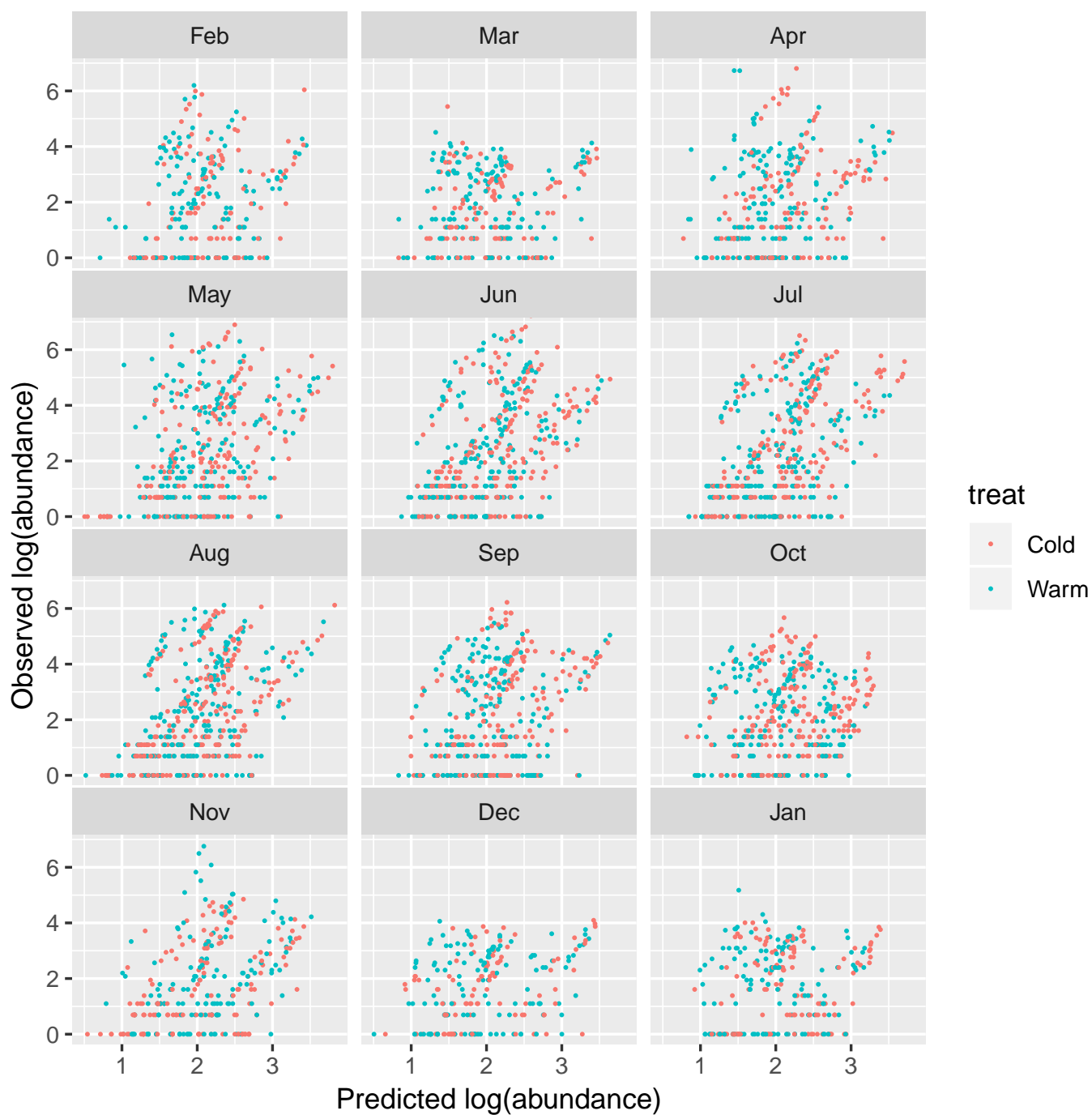
Figure 4: **The observed value and predicted value of the allometric scaling model.** The red points in graph represent samples of warming condition and the blue points represent samples of ambient condition.

3 Plante-Downing model

The coefficients in Plante-Downing model of different conditions were listed in Table 3. According to the
plotting of the observed value against predicted value of $log(abundance)$ as shown in Figure 5, Plante-
Downing model fitted expectedly well into the model.

	Intercept	log(bodymass)	log(max.bodymass)	temperature
Feb	1.82415816109381	1.06952012596034	−1.02126289890586	−0.00134713249771387
Mar	1.94193010875633	1.0628880098179	−1.03243272720307	−0.0123673150811795
Apr	1.77721816000063	1.08557343820203	−1.05928248927854	−0.000621584922526176
May	1.53452157903706	1.08246932372983	−1.05117069198733	0.0124890386285008
Jun	0.166314433524029	1.07664558360067	−1.04676280840836	0.0720526700071134
Jul	0.619054791987107	1.05961625785255	−1.03846833020904	0.0494897852421087
Aug	1.03104709841322	1.06372176328331	−1.03022328178242	0.0352710005508637
Sep	2.06355380161005	1.09657817539538	−1.05949767185426	−0.0136368810274392
Oct	2.1076034331867	1.08138799751452	−1.05366235591937	−0.0201447631511348
Nov	2.28055812148817	1.12038497243504	−1.06198544642223	−0.0183586468839484
Dec	1.7011699836853	1.08454099466958	−1.0783886999298	NA
Jan	1.71379889579386	1.0873718763242	−1.05778868106958	0.00281032650217881

	Intercept	log(bodymass)	log(max.bodymass)	temperature
Feb	1.62085155233827	1.07626502942791	−1.05269614061308	0.00476433770575302
Mar	1.68994542267236	1.09199051175804	−1.05439290358038	0.0144939873229885
Apr	1.90904944622679	1.09260260093098	−1.06167449550784	−0.0124154413922979
May	1.67329272331735	1.0683498151909	−1.02607563514963	0.0070917111388728
Jun	1.95969863258866	1.05889617045759	−1.01319947123621	−0.006722252611165
Jul	1.99908024771756	1.04437524728099	−1.00685493181354	−0.0102106344073382
Aug	1.4765677494378	1.05005364306172	−1.00367460320311	0.0206318920701787
Sep	2.2369372225962	1.10134167641945	−1.04891669060001	−0.0241200657530604
Oct	2.129804799609	1.06856160746788	−1.03087947310114	−0.0222853653085298
Nov	1.88898303678631	1.05974671799246	−1.03828178484637	−0.0172202461667939
Dec	1.83555745244096	1.10380031846922	−1.07104149528508	−0.00620694891876475
Jan	1.74003021459663	1.09138072521767	−1.06628116232399	−0.00375469699641393

Table 3: **The coefficients of the Plante-Downing model.** The upper table is the coefficients under warming condition. The lower table is the coefficients under ambient condition.
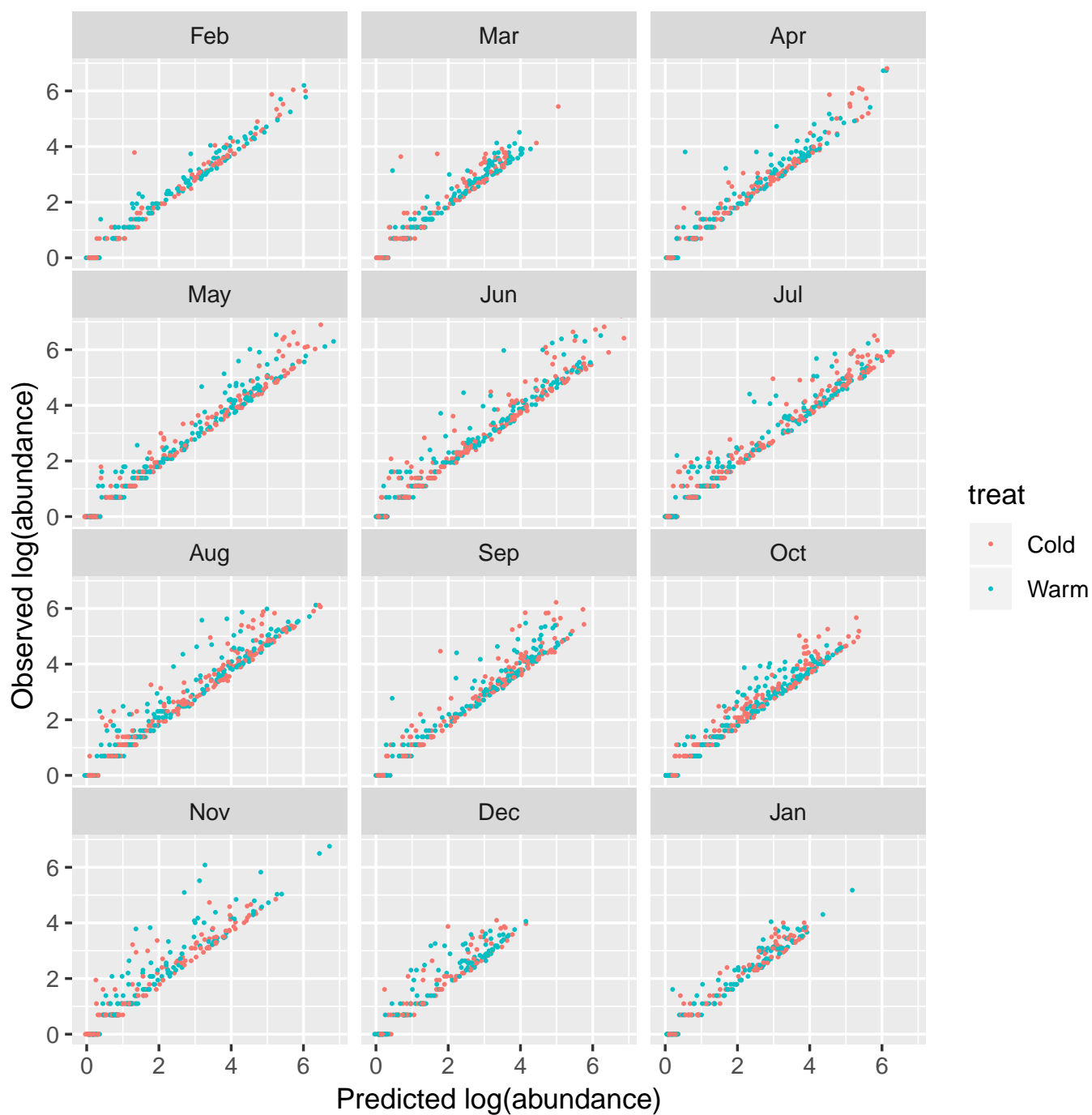
Figure 5: **The observed value and predicted value of the Plante-Downing model.** The red dots in graph represent samples of warming condition and the blue dots represent samples of ambient condition.

4 Model selection

Generally, models with relatively low AIC are regareded as better-fitted models. Based on the AIC results of both models shown in Table 4, it is clear that Plante-Downing model is a better model to explain the interrelations between abundance and bodymass of freshwater microbial communities.

	warm AIC	cold AIC		warm AIC	cold AIC
Feb	107.620243361704	110.560093094507	*Feb*	−364.138445351006	−284.300465088297
Mar	66.5507995567171	63.2273736520536	*Mar*	−331.556131025574	−216.94459098115
Apr	142.298286754982	152.075630235894	*Apr*	−283.649660015521	−352.582443605816
May	217.808296086022	252.026150740786	*May*	−377.068098835864	−463.322943139075
Jun	242.036258277027	272.501994388019	*Jun*	−356.208163932737	−421.438270025188
Jul	241.409169049025	284.483080914331	*Jul*	−389.024530307358	−433.441269651944
Aug	246.919363101043	264.974862451197	*Aug*	−362.833257707643	−446.898262057935
Sep	205.342002630937	217.440679439548	*Sep*	−427.745985911502	−357.942906962413
Oct	150.950815143152	151.880418408462	*Oct*	−545.686283377122	−556.344167000338
Nov	143.094063509075	143.31099850421	*Nov*	−160.119351324837	−315.726019212047
Dec	70.9916935552908	50.6581406030575	*Dec*	−237.511239218527	−192.590593124469
Jan	60.2123774004006	77.5248352984576	*Jan*	−308.598396256712	−342.835313257915

Table 4: **The AIC value of the two model under both warming and ambient condition.** The table on the left is the AIC value of the allometric scaling model.The table on the right is the AIC value of Plante-Downing model.

Discussion

From the above analysis, it is clearly concluded that Plante-Downing model is the better choice for freshwater microbial communities data analysis than the allometric scaling model as it showed a more firm scatterplot fitting, while the widely used allometric scaling model still provided us with a reference and some feedback. The lack of performance of allometric scaling model in AIC largely attributed to the small abundance in many sampled species. It could be observed in Figure 3 that a certain amount of scattering points lying on the line $y = 0$, indicated that the abundance value of these sampled species in that month is 1.

Although the Plante-Downing model revealed a good performance in AIC, it is necessary to realize that this empirical model was very likely to have a good fit for the data in this case as the biomass here

was calculated by the multiplication of bodymass and abundance. Assuming A as abundance and B as bodymass, subsituting $\overline{B}$ with $B_{mean} * A$ into the formula 2, then we can discover that $\log(A) = a + b\log(B_{mean} * A) - c\log(B_{max}) + dT$. This formula can be transformed to $\log(A) = \log((B_{mean} * A)^b / \log(B_{max})^c) + dT + a$. Due to the appearance of abundance on both sides of the equation, the model is available to perform a great fit by adjustment of constant values b and c. For further optimizing the Plante-Downing model of this case, it is better to use projection to calculate the population of freshwater microbial communities in the following months, and then the abundance of each month are able to link sequentially.

In addition, the exclusion of some abnormal data might affect the model fitting results. Those abnormal data have been deleted because the microbial individuals were recorded as spherical shape whereas the length and width recorded was quite different. For example, the length and width of taxa recorded having a spherical shape is 1200 μm and 21 μm. The total number of data deletion was 313 and it is necessary to remain the doubts of unknown influences to the model fitting results. Thoses deleted individuals might play important roles in the freshwater microbial communities such as a predator. For further detail studies, it is preferable to implement advanced data analysis algorithms rather than simply omitting these abnormal data.

Conclusion

In conclusion, the Plante-Downing model performed better than the allometric scaling model in terms of AIC. The results of this article supported Yvon's statement that freshwater microbial communities biomass decreases in warming condition. This study also validated Yvon's statement that the gradient of linear litting in warming condition was steeper, implicating freshwater microbial communities at a relatively lower growth rate of abundance under warming condition.

14

References

[1] Ken Aho, DeWayne Derryberry, and Teri Peterson. Model selection for ecologists: the worldviews of aic and bic. *Ecology*, 95(3):631–636, 2014.

[2] NW Arnell and NS Reynard. The effects of climate change due to global warming on river flows in great britain. *Journal of hydrology*, 183(3-4):397–424, 1996.

[3] Stephen R Carpenter, Stuart G Fisher, Nancy B Grimm, and James F Kitchell. Global change and freshwater ecosystems. *Annual Review of Ecology and Systematics*, 23(1):119–139, 1992.

[4] Helmut Hillebrand, Claus-Dieter Dürselen, David Kirschtel, Utsa Pollingher, and Tamar Zohary. Biovolume calculation for pelagic and benthic microalgae. *Journal of phycology*, 35(2):403–424, 1999.

[5] Julian Sorell Huxley. Relative growth and form transformation. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, 137(889):465–469, 1950.

[6] IPCC. Global warming of 1.5 c, 2018.

[7] Jerald B Johnson and Kristian S Omland. Model selection in ecology and evolution. *Trends in ecology & evolution*, 19(2):101–108, 2004.

[8] Robert Henry Peters and Karen Wassenberg. The effect of body size on animal abundance. *Oecologia*, 60(1):89–96, 1983.

[9] Celine Plante and John A Downing. Production of freshwater invertebrate populations in lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 46(9):1489–1498, 1989.

[10] Julia Reiss and Jenny M Schmid-araya. Existing in plenty: abundance, biomass and diversity of ciliates and meiofauna in small streams. *Freshwater Biology*, 53(4):652–668, 2008.

[11] Jennifer A Sheridan and David Bickford. Shrinking body size as an ecological response to climate change. *Nature climate change*, 1(8):401, 2011.

[12] Rebecca Stewart. Global warming in freshwaters:implications for the microbial-meiofaunal loop. 2012.

[13] Eric-Jan Wagenmakers and Simon Farrell. Aic model selection using akaike weights. *Psychonomic bulletin & review*, 11(1):192–196, 2004.

[14] Richard John Walters and Mark Hassall. The temperature-size rule in ectotherms: may a general explanation exist after all? *The American Naturalist*, 167(4):510–523, 2006.

[15] S Weart. The discovery of global warming: The public and climate change (cont.–since 1980), 2009.

[16] Gabriel Yvon-Durocher, José M Montoya, Mark Trimmer, and GUY Woodward. Warming alters the size spectrum and shifts the distribution of biomass in freshwater ecosystems. *Global change biology*, 17(4):1681–1694, 2011.