Faculty Authored Articles

2018

# Modeling Mayfly Nymph Length Distribution and Population Dynamics Across a Gradient of Stream Temperatures and Stream Types

Jeremy Anthony
*Augsburg University*

Jennifer Baccam
*Augsburg University*

Imanuel Bier
*Augsburg University*

Emily Gregg
*Augsburg University*

Leif Halverson
*Augsburg University*

***See next page for additional authors***

Follow this and additional works at: https://idun.augsburg.edu/faculty_scholarship

Part of the Applied Statistics Commons, Ordinary Differential Equations and Applied Dynamics Commons, and the Other Environmental Sciences Commons

Recommended Citation

Anthony, Jeremy; Baccam, Jennifer; Bier, Imanuel; Gregg, Emily; Halverson, Leif; Mulcahy, Ryan; Okanla, Emmanuel; Osman, Samira A.; Pancoast, Adam R.; Schultz, Kevin C.; Sushko, Alex; Vorarath, Jennifer; Vue, Yia; Wagner, Austin; Schilling, Emily Gaenzle; and Zobitz, John M. (2018) "Modeling Mayfly Nymph Length Distribution and Population Dynamics Across a Gradient of Stream Temperatures and Stream Types," Spora: A Journal of Biomathematics: Vol. 4 : Iss.1, 1 — 1.

**Authors**

Jeremy Anthony, Jennifer Baccam, Imanuel Bier, Emily Gregg, Leif Halverson, Ryan Mulcahy, Emmanuel Okanla, Samira A. Osman, Adam R. Pancoast, Kevin C. Schultz, Alex Sushko, Jennifer Vorarath, Yia Vue, Austin Wagner, Emily Gaenzle Schilling, and John Zobitz

# Modeling Mayfly Nymph Length Distribution and Population Dynamics Across a Gradient of Stream Temperatures and Stream Types

**Authors**

Jeremy Anthony, Jennifer Baccam, Imanuel Bier, Emily Gregg, Leif Halverson, Ryan Mulcahy, Emmanuel Okanla, Samira A. Osman, Adam R. Pancoast, Kevin C. Schultz, Alex Sushko, Jennifer Vorarath, Yia Vue, Austin Wagner, Emily Gaenzle Schilling, and John M. Zobitz

# Modeling Mayfly Nymph Length Distribution and Population Dynamics Across a Gradient of Stream Temperatures and Stream Types

Jeremy Anthony[1], Jennifer Baccam[1], Imanuel Bier[1], Emily Gregg[1], Leif Halverson[1], Ryan Mulcahy[1], Emmanuel Okanla[1], Samira A. Osman[1], Adam R. Pancoast[1], Kevin C. Schultz[1], Alex Sushko[1], Jennifer Vorarath[1], Yia Vue[1], Austin Wagner[1], Emily Gaenzle Schilling[2], John M. Zobitz[3],*

*Correspondence:
Prof. John Zobitz,
Department of Mathematics,
Statistics, & Computer
Science, Augsburg University,
Minneapolis, Minnesota
55454
zobitz@augsburg.edu

**Abstract**

We analyze a process-based temperature model for the length distribution and population over time of mayfly nymphs. Model parameters are estimated using a Markov Chain Monte Carlo parameter estimation method utilizing length distribution data at five different stream sites. Two different models (a standard exponential model and a modified Weibull model) of mayfly mortality are evaluated, where in both cases mayfly length growth is a function of stream temperature. Based on model-data comparisons to the modeled length distribution and the Bayesian Information Criterion, we found that approaches that length distribution data can reliably estimate 2–3 model parameters. Future model development could include additional factors include such as upstream environmental factors, abiotic conditions, interspecific competition, predation, or stream salinity. Outputs of this model could be applied to predict mayfly emergence across a geographic domain or to forecast mayfly population responses to climate change.

**Keywords:** mayfly, Markov Chain Monte Carlo, population dynamics, data assimilation

Mayflies (order Ephemeroptera) are one of several insect taxa that are aquatic as juveniles and terrestrial as adults. Over the course of their lifespan mayflies are a key contributor to aquatic food webs [12]. As juveniles, mayflies are benthic organisms associated with stream beds. The juvenile life history stage may range from days to years, during which time their final body length ranges from 3 to 30 millimeters depending on species [6]. A contributing factor to mayfly development across a range of species and aquatic ecosystems is water temperature [9, 12, 20, 26, 30]. One approach to understanding this response to temperature, especially in relation to development time to adult emergence, is through analysis of a process model parameterized by stream temperature.

Typically nymph length development is modeled with a threshold function that assumes growth occurs in a fixed temperature range [20]. This response predicts any shifts to stream temperature outside this range may impact both the development and timing of emergence, especially for cold-water adapted species [10, 29]. Based on current and predicted patterns on global greenhouse gas emissions, stream water temperatures in the next century could increase on average 1.0 °C, with expected variability in the magnitude of warming on a regional scale [27]. Consequently, any shift in water temperature has the potential to impact local aquatic food webs [15, 28].

A standard way to quantify mayfly nymph length is the size frequency method, which reports probability length distributions of mayfly nymphs sampled monthly, typically reported over a year [1]. This method has been previously applied to investigate life history and development of different mayfly species adapted to a particular stream [2, 5, 7, 8, 13, 14, 16, 29]. Although aggregation and synthesis of length distribution studies have been rare (with the exception of Clifford [6] and Huryn and Wallace [12]), monthly measurements of nymph length distribution, in conjunction with stream temperature, could help determine key parameters that affect any mathematical model of nymph population dynamics [20].

The objective of this study is to parameterize and compare models of mayfly nymph development utilizing a range of published life history studies. We develop models that (1) assume nymph length development is solely a function of stream temperature and (2) consider a constant or time-varying mortality function for individual nymphs. Model results are compared against life his-

[1]Augsburg University, Minneapolis, MN, [2]Department of Biology, Augsburg University, Minneapolis, MN, [3]Department of Mathematics, Statistics, & Computer Science, Augsburg University, Minneapolis, MN

tory studies of nymph length distribution. To simplify our approach, we only consider mayfly species with annual life cycles. We make two key assumptions in this study: (1) application to the same model to similar life history studies provides information as a baseline for how different native species respond and (2) parameters for a given model are applicable in a broad range of contexts and climates, even though each separate mayfly species might be bioadapted to its own climate.

# 1 Methods

This study utilizes published data from five mayfly life history studies that report (1) length density distributions and (2) stream temperatures over a period of several months. We consider five different species and streams described in Table 1. To generate data from published temperature graphs we utilize the open-software program Engauge (http://markummitchell.github.io/engauge-digitizer/), interpolated to daily values using piecewise linear interpolation. Monthly mayfly length distribution was determined via measuring tools found with portable document format reader software such as Preview or Reader. We then subsequently convert these measurements to a probability density distribution. We use the program R and RStudio [22, 23] for all software analyses. All studies report mayfly body lengths, with the exception of Winterbourn et al. [29], which measured head width. In that case, for simplicity we assume that the probability distribution of head width length is similar to the body length distribution.

## 1.1 Mayfly population model

The model we utilize for mayfly growth is similar to the one described previously in Zobitz et al. [31]. This model has two dynamic variables of interest: $M$, the total population of mayfly nymph, and $l_i(t)$, the length of an individual mayfly $i$ at time $t$, where time is measured in months. For this model, the length rate of change of a mayfly nymph is proportional to stream temperature when the daily temperature $T$ is between $T_L$ or $T_L + R$, where $R$ is the thermal range for growth, according to the following equation:

$$\frac{dl_i}{dt} = \begin{cases} \alpha & \text{for } T_L \leq T < T_L + R, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

where $\alpha$ (mm day$^{-1}$) is the length growth rate of mayflies. When there are $D$ days in a month when the daily temperature is between this optimum range each mayfly will accumulate $\alpha D$ units of length.

We assume that the population of mayflies $M$ follows the following exponential differential equation:

$$\frac{dM}{dt} = -\delta(t)M, \quad (2)$$

where $\delta(t)$ (% mayflies month$^{-1}$) is a time-dependent mortality of mayfly nymph. The basic solution to this differential equation is given by

$$M(t) = M_0 \exp\left(-\int_0^t \delta(s)\, ds\right), \quad (3)$$

where $M_0$ is the initial population of mayflies that follows a prescribed length distribution. For all streams and simulations we assumed $M_0 = 1000$. In our analyses we consider two formulations for $\delta(t)$:

- When $\delta(t)$ is constant Equation 3 is the exponential model of population growth, or

$$M(t) = M_0 \exp(-\delta t). \quad (4)$$

 We will denote this model as the **Standard model**.

- If we assume nymph mortality increases over time, one possible formulation is $\delta(t) = \delta \cdot \left(\frac{t}{\tau}\right)^p$, where $\delta$ is a constant (as in the Standard model), $p$ is a shape parameter that is greater than zero, and $\tau$ is the total time (measured in months). This model is an example of a Weibull population model [17]. The solution for Equation 3 is

$$M(t) = M_0 \exp\left(-\frac{\delta \cdot \tau^{p+1}}{p+1}\left(\frac{t}{\tau}\right)^{p+1}\right)$$
$$= M_0 \exp\left(-\frac{\delta \cdot t^{p+1}}{p+1}\right). \quad (5)$$

 We will denote this model as the **Weibull model**.

In both cases the model is implemented on a monthly timestep using Euler's method. At each month mayfly mortality is implemented by removing no more than a percentage of $\delta(t)$ mayflies each month. A summary of each model and the parameters used in both is shown in Table 2.

## 1.2 Parameter estimation method

To estimate parameters we minimize the difference between the modeled and measured mayfly length density across each month for each site. Autocorrelation tests on the mean mayfly length did not detect a significant presence of autocorrelation in the measurements (results not shown). Of all five sites, only consecutive monthly

Table 1: Description of mayfly life history studies analyzed.

| Site | Notes |
| --- | --- |
| Prosiek Stream, Slovakia [2] | Located in the West Carpathians in Northern Slovakia (49°9'N, 19°29'E), 705 meters above sea level. For this study we analyzed length distribution data for *Rhithrogena carpatoalpina*. |
| | Bottová et al. [2] report stream temperatures of 8 °C from May to August, and 7 °C from September to April. We model this annual pattern with constant decrease of 0.23 °C per day between August 15 to September 15. Following construction of that piecewise linear function we account for temperature variations by adding random noise to the daily temperature sampled from a normal distribution with mean zero and standard deviation 0.1 °C. |
| Mobile River, USA [5] | Located between Alabama, northwestern Georgia, and eastern Mississippi. Chadwick and Feminella [5] collected samples of the mayfly species *Hexagenia limbata* across several sites ranging from a latitude of 30°40' to 31°N, 88°W from October 1995 to November 1996. The recorded stream temperatures ranged from 5 °C to 30 °C. |
| Sutton Stream, New Zealand [13] | A high-elevation tributary of the Taieri River on the South Island of New Zealand (45°36'S, 169°54'E). The life history of populations of the leptophlebiid mayfly *Deleatidium* (*lillii* group) located in during the period extending from October 31, 1991 to October 31 1992. Over the observation period stream temperatures ranged from 0 °C to 17.4 °C. |
| Gapyeong Stream, South Korea [14, 16] | Located in South Korea (37°57'N, 127°15'E). Samples of *Ephemera orientalis McLachlan*, a common burrowing mayfly distributed in temperate East Asia were collected March 1998 to June 1999. The recorded stream temperatures during the study period range from 3 °C to 25 °C. |
| Rob Roy Stream, New Zealand [29] | Rob Roy Stream (44°26'S, 168°43'E) is a south-facing tributary of the Matukituki River in the southwest of the South Island, New Zealand. Head width arval samples of the mayfly *Deleatidium cornutum* (subgenus *Penniketellum*) were analyzed from November 2004 to October 2005. Mayflies in this stream are cold adapted; average water temperature over the period was 2.1 °C, ranging from 1.2 °C to 3.7 °C. |
| | Winterbourn et al. [29] did not report a timeseries of temperature data. To generate a representative graph of stream water temperature, we assumed that the monthly water temperature is proportional to the air temperature found from local climatological records (`http://www.yr.no/place/New_Zealand/Otago/Rob_Roy_Peak/statistics.html`). We then interpolated the pattern of the monthly air temperature to a daily value, adding independent random noise of mean zero and standard deviation equality to the square root of the reported range in temperatures (2.5 °C). By constructing a representative stream temperature timeseries in this way, the average of that data approximately equaled the published average mean of 2.1°. |

Table 2: Description of model parameters. $\mathcal{S}$: Standard model (Equation 4) $\mathcal{W}$: Weibull model (Equation 5)

| Symbol | Units | Description | Model | Prior Range |
|:---:|:---:|:---:|:---:|:---:|
| $T_L$ | °C | Minimum temperature for growth | $\mathcal{S}, \mathcal{W}$ | 0–20 |
| $R$ | °C | Thermal range for mayfly growth | $\mathcal{S}, \mathcal{W}$ | 0–30 |
| $\alpha$ | mm day$^{-1}$ | Mayfly length growth rate | $\mathcal{S}, \mathcal{W}$ | 0–1 |
| $\delta$ | % mayflies month$^{-1}$ | Mayfly population death rate | $\mathcal{S}, \mathcal{W}$ | 0–1 |
| $p$ | no units | Shape parameter | $\mathcal{W}$ | 0–5 |

measurements of mean mayfly length at Rob Roy stream were significantly correlated ($p < 0.02$). We assume that each measurement is independent, identically distributed, so we have the following log likelihood function (denoted as $\mathcal{LL}$) in Equation 6:

$$\mathcal{LL}(\vec{v}) = -N \ln(\sqrt{2\pi}) - \sum_i^N \ln(\sigma_i) - \sum_{i=1}^N \frac{(x_i - \eta_i(\vec{v}))^2}{2\sigma_i^2},$$
$$(6)$$

where $x_i$ is a measured value, $\eta_i$ is a modeled value, $\vec{v}$ the vector of estimated parameters, $N$ the number of data points, and $\sigma_i$ the uncertainty on the measurements. For this study $\sigma_i$ is the uncertainty in the digitization of the distribution measurements. Preliminary investigations measuring the length distribution found it computationally feasible to measure the length distribution curve to high precision, which effectively forces Equation 6 to have zero weight on the measured values. To modify this effect we treated $\sigma_i$ as a nuisance parameter and set $\sigma_i$ equal to the value that maximizes log-likelihood, which can be shown to equal the sample standard deviation [3].

For each parameter we initially prescribe a uniform probability distribution over a selected range shown in Table 2. For convenience we assume that the range of each parameter is the same across each of the sites. Numerical bounds for the prior parameter values were selected based on reasonableness (e.g. mayflies cannot grow below 0 °C) or previous studies utilizing this model [31]. The shape parameter $p$ was selected between zero and five because if $p$ is much greater than unity if effectively forces $\delta(t) = 0$, causing the mayfly population to remain constant.

To estimate parameters we apply a variation of the Metropolis algorithm [11]. The Metropolis algorithm is an iterative search through the multidimensional parameter space. At each iteration the parameter estimation routine randomly modifies a particular parameter to a proposed new value, evaluates the log-likelihood, and compares it to the current log-likelihood. We accept the proposed parameter set if the log-likelihood increases. If the proposed parameter set does not increase the joint log-likelihood it is still accepted with a probability proportional to the difference of the log-likelihoods. With this setup approximately 30% of parameter proposals are accepted.

To avoid convergence to a local optimum, we first run seven independent chains until obtaining the prescribed acceptance rate of 30% or by reaching a finite number of iterations. The chain that has the highest log likelihood is selected for parameter estimation. The parameter estimation chain has 150,000 iterations, of which the first 70,000 iterations are discarded as the "burn-in" period when the parameter estimation is converging to the optimum. Final parameter estimates and analysis are computed on the last 80,000 iterations.

Two different parameter estimates were generated for each of the two models studied (Standard and Weibull model). In the first approach, we estimated 4 or 5 parameters ($T_L$, $R$, $\alpha$, $\delta$, $p$) depending on if it was the Standard or Weibull model. Additionally for each stream we fixed the values of $T_L$ and $R$ so these values fell between the reported stream temperatures for each site. This second approach allowed us to investigate the effect of reducing the parameter dimension on the parameter estimation.

This sampling strategy allows several post-assimilation analyses. First summary statistics (median and confidence intervals) for each parameter sets are computed. To assess goodness of fit and determination of the best approximating model for each site, we compare modeled mayfly length to measured values. The Bayesian Information Criterion is applied to determine the best approximating model [4, 24]:

$$BIC = -2 \cdot \mathcal{LL}_{max} + P \cdot \ln(N), \qquad (7)$$

where $\mathcal{LL}_{max}$ is the maximum log-likelihood, $P$ the number of estimated parameters, and $N$ is the number of data points. The best approximating model is the one with the lowest $BIC$. We also generate population forecasts by first randomly thinning the set of accepted parameter values by 50% for computational efficiency, next evaluating the mayfly model for each of the remaining sets of

parameters. and then reporting results at each timestep as an ensemble average (median plus 95% centered confidence interval).

## 2   Results

MCMC parameter estimates for $\alpha$, $\delta$, and $p$ were well-constrained by the data whereas the parameters $T_L$ and $R$ were not as constrained from the data (Figure 1). In both cases, common parameters across the Standard and Weibull models were proximate, but not identical, in value. In the case when $T_L$ and $R$ were fixed parameter estimates for $\alpha$ were close four of the five sites. Generally speaking the data contained information that was able to reliably estimate two to three parameters at each site.

We further analyzed the impact of the parameter estimates of $T_L$ and $R$ in relation to the stream temperature data. For two out of the five sites (Mobile River and Gapyeong Stream) the median values of $T_L$ and $R$ fall within the measured ranges stream temperature (Figure 2).

Even though the parameter estimates were well constrained, further examination of the modeled temperature ranges compared to the measured temperature utilized shows the effect of mayfly mortality on the length distribution. All of the models were able to reasonably capture the measured mayfly length distribution at each site when compared across each month (Figures 3 and 4, results not shown when $T_L$ and $R$ are held fixed because they were similar). Because a plot of measured versus modeled values contains uncertainties in the measurement, we calculated regression slope using geometric mean regression [25]. A bootstrap estimate [17] of the 95% confidence interval for the regression slope included unity for three of the five sites (Figure 5 and Table 3).

Population forecasts with both the Standard and Weibull model suggest a decreasing population over time (Figure 6). Generally speaking the Standard model displayed less uncertainty in modeled population values than the Weibull model, however the Weibull model had a larger survivorship of nymph than the Standard model. For both of these models there was no appreciable difference between the population forecasts results corresponding to the same model when $T_L$ and $R$ were fixed (results not shown).

## 3   Discussion

Our objective of this study was to parameterize and evaluate a mayfly population model that considered the distribution of individual nymph length and the mayfly population. We compared two different models of nymph

mortality across several different ecosystem types and latitudes. For several sites, population level parameters ($\alpha$, $\delta$, $T_L$ and $R$) were consistent and comparable (Figure 1).

Several observational studies have posited temperature is a key driver of nymph length development [12, 20, 26, 30]. Our results comparing estimated values of $T_L$ and $R$ to when these two parameters are fixed highlight (1) the mechanistic meaning of these parameters under the assumption of temperature dependent length growth and (2) the importance of these values falling within the reported stream temperatures. Figure 2 shows that temperature parameter estimates of $T_L$ and $R$ were reasonable (meaning the final parameter estimates were within the range or stream temperatures) for only two of the five sites studied. Parameter estimates of $T_L$ and $R$ outside the measured range of temperatures (as in the case of Prosiek, Sutton, and Rob Roy Streams) would imply that mayfly length distribution would be constant in time. In contrast, measured length distributions for these sites show variation over time (Figures 3 and 4). If we set the value of $T_L$ and $R$ to be within the reported stream temperature range, either the estimated value of the growth rate $\alpha$ tends towards zero (meaning the length distribution is stationary) or nymph mortality $\delta$ tends to unity (meaning populations decline rapidly), see Figure 1. In both of these cases $\alpha$ and $\delta$ serve as a tempering parameter to prevent skewing the length distribution to larger mayflies. From these results we conclude that $T_L$ and $R$ cannot be independently estimated from length distribution data alone and recommend specifying them for each stream.

For the two models and data considered in this study using the $BIC$, in Table 3, three of the five sites suggest that a non-constant mortality (as in the Weibull model) is preferred compared to a constant mortality. Related to this conclusion, four of the five sites suggest that $T_L$ and $R$ should be fixed parameters. Future work is needed to determine if this result would hold across other stream types and nymphs, as three ecological factors favor a non-constant mortality function. First, mayfly nymphs may be affected by site specific environmental and ecological factors such as upstream environmental factors, abiotic conditions, interspecfic competition, predation, or stream salinity [7, 8]. All of these factors may suggest that future studies should investigate a length dependence on mortality in addition to a temporal dependence on mortality. Second, while the BIC indicates a slight preference for the Standard model, the population forecasts (Figure 6) suggest a steeper decline in mayfly survivorship over time than forecasts with the Weibull model. Additional data, such as population counts over time, could be incorporated into the parameter estimation to further elucidate the best approximating model from the data. Third, the current analysis opted for a model that started
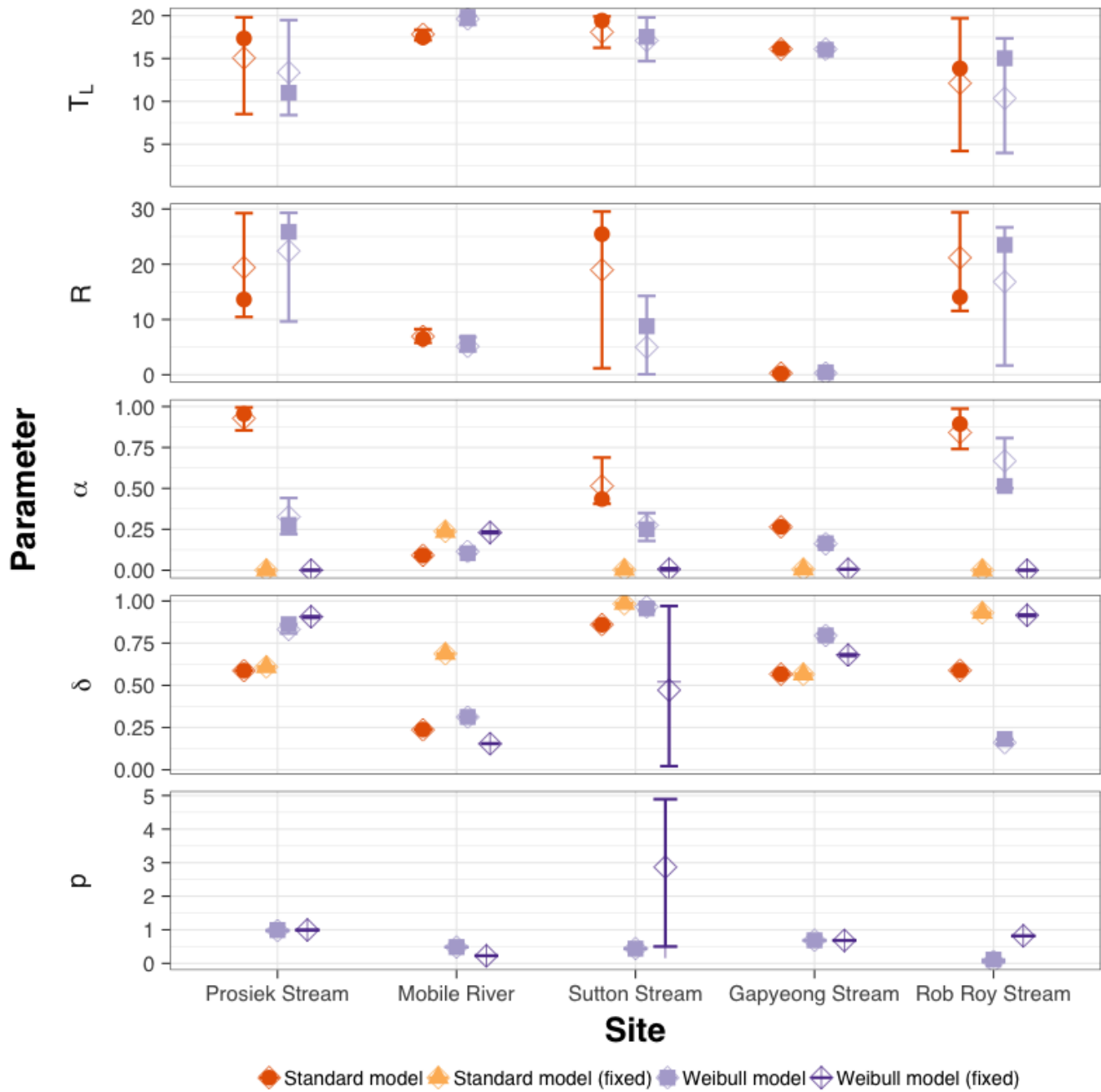
Figure 1: Parameter estimation results for the two different models considered (Standard model, Equation 4 and Weibull model, Equation 5) across the range of sites as well as when $T_L$ and $R$ were assigned values for each site (Standard model (fixed) and Weibull model (fixed)). The whiskers for each parameter represent the 95% confidence interval of the parameter distribution, the empty diamond the median, and the filled triangle the parameter that maximized the log-likelihood.
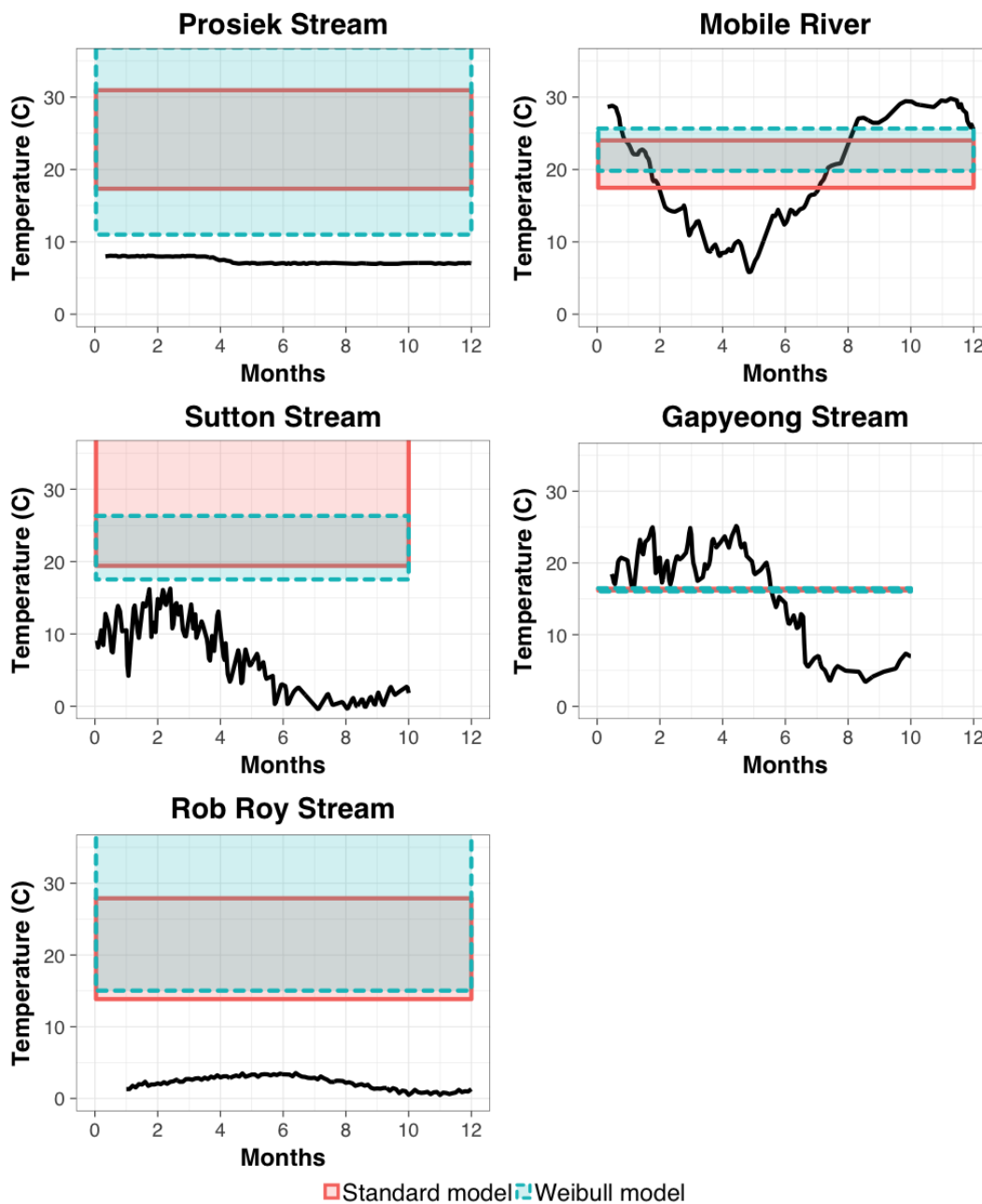
Figure 2: Parameter estimation for $T_L$ and $R$ for the two different models considered (Standard model, Equation 4 and Weibull model, Equation 5) across the range of sites along with the measured stream temperatures (solid line). Values for $T_L$ and $R$ are selected from the median parameter estimate of each site in Figure 1. The shading indicates the fixed values of $T_L$ and $R$ for each stream for the Standard model. In the case of Standard model (fixed) and Weibull model (fixed), values of $T_L$ and $R$ are provided in Table 2.
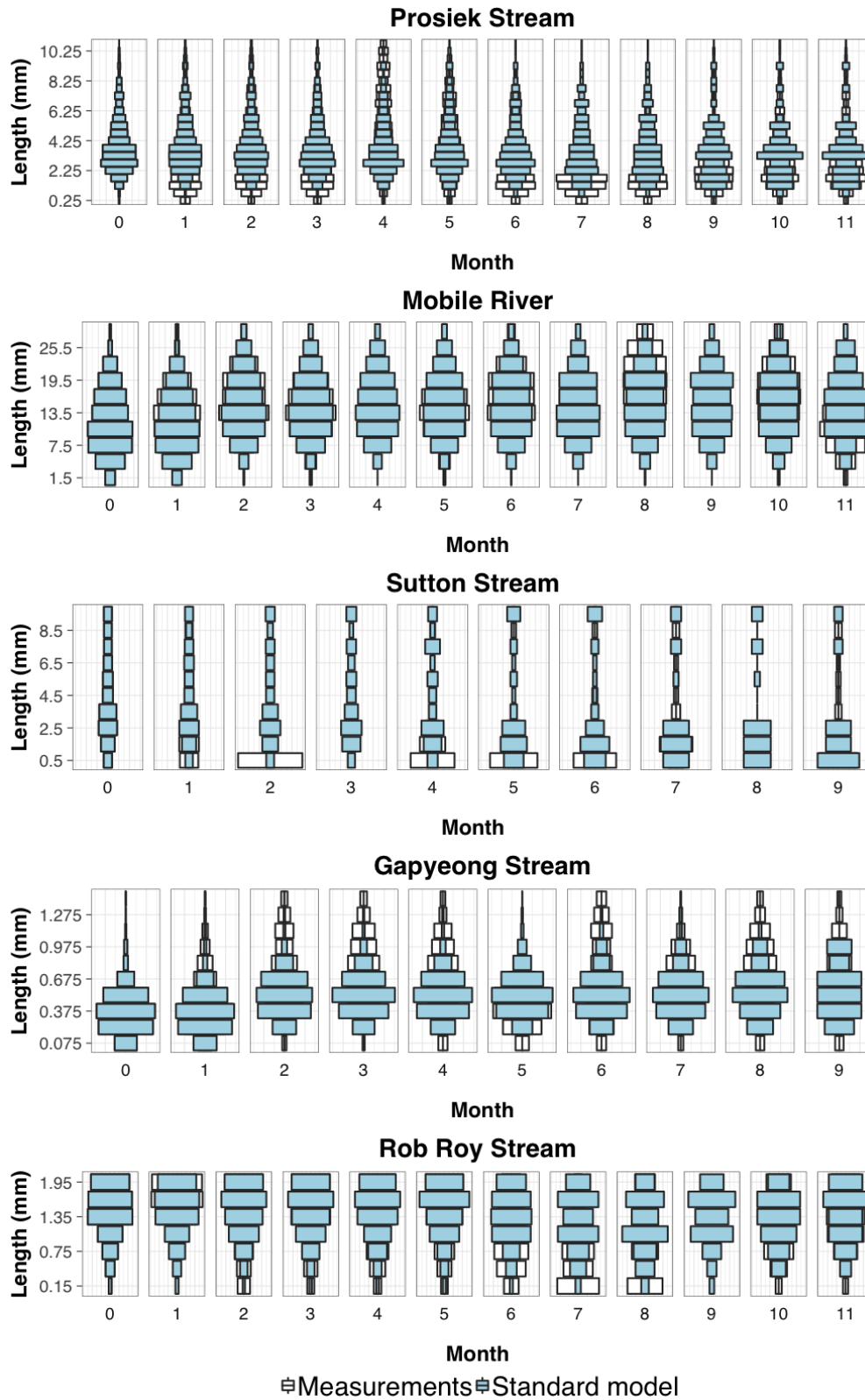
Figure 3: Comparison of measured and modeled length densities over time for each site. Model results shown are for the Standard model (Equation 4).
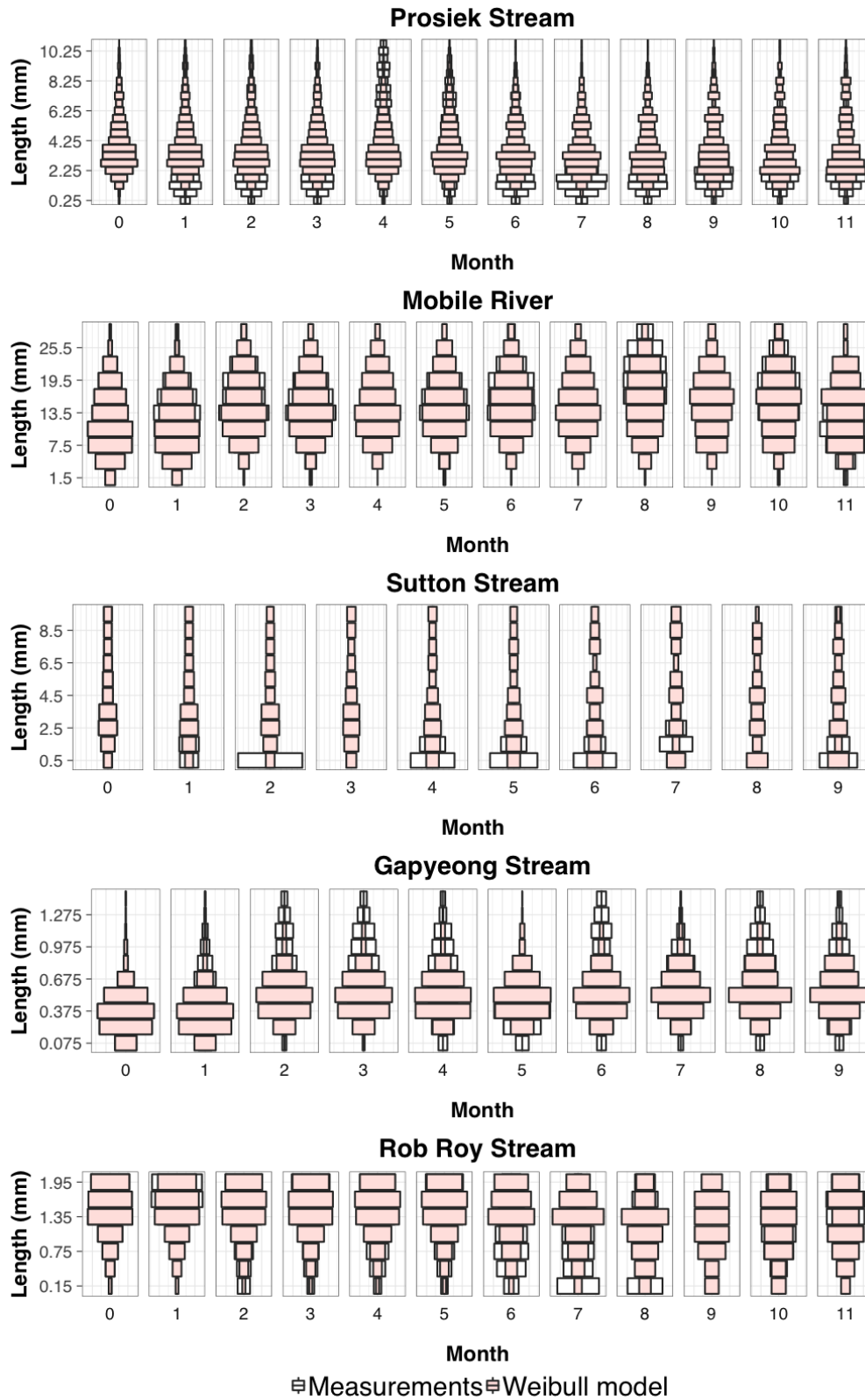
Figure 4: Comparison of measured and modeled length densities over time for each site. Model results shown are for the Weibull model (Equation 5).
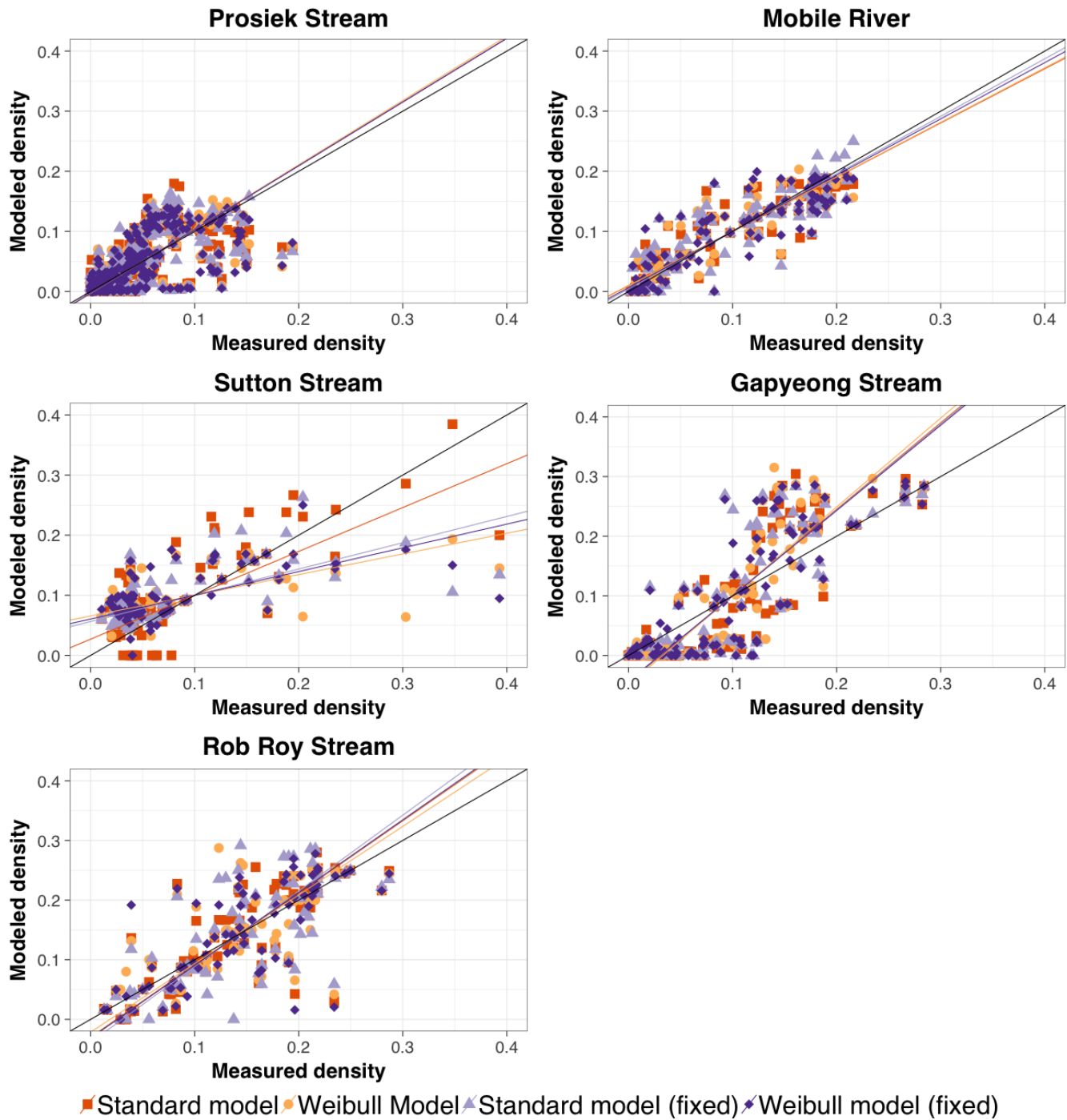
Figure 5: Comparison of measured and modeled length densities across the different sites for the two different models considered (Standard model, Equation 4 and Weibull model, Equation 5), as well as when $T_L$ and $R$ are fixed or estimated parameters. The 1:1 line is included for comparison.

Table 3: Information criteria and model-data fits by site and model. $N$: the number of measurements used in the parameter estimation. $P$: number of parameters estimated. In the case of Standard model (fixed) and Weibull model (fixed), values of $T_L$ and $R$ are provided in Table 2. $m$: slope of linear regression from Figure 5, calculated via geometric mean regression, followed with the 95% confidence interval calculated via bootstrap sampling. $R^2$: coefficient of determination. $\mathcal{LL}$: Maximum log-likelihood. $BIC$: Bayesian information criterion from Equation 7. The best approximating model is the one with the lowest $BIC$.

| Site / Model | $N$ | $P$ | $m$ | $R^2$ | $\mathcal{LL}$ | $BIC$ |
|---|---|---|---|---|---|---|
| **Prosiek Stream** | | | | | | |
| Standard model | 264 | 4 | 1.1 (0.9, 1.2) | 0.45 | 512 | −1001 |
| Weibull model | 264 | 5 | 1.1 (0.9, 1.2) | 0.41 | 497 | −966 |
| Standard model (fixed) | 264 | 2 | 1.1 (0.9, 1.2) | 0.43 | 511 | **−1011** |
| Weibull model (fixed) | 264 | 3 | 1.1 (0.9, 1.2) | 0.41 | 506 | −995 |
| **Mobile River** | | | | | | |
| Standard model | 90 | 4 | 0.90 (0.8, 1.0) | 0.80 | 178 | −338 |
| Weibull model | 90 | 5 | 0.91 (0.8, 1.0) | 0.83 | 183 | **−343** |
| Standard model (fixed) | 90 | 2 | 0.96 (0.9, 1.1) | 0.74 | 170 | −321 |
| Weibull model (fixed) | 90 | 3 | 0.94 (0.8, 1.0) | 0.76 | 173 | −333 |
| **Sutton Stream** | | | | | | |
| Standard model | 80 | 4 | 0.73 (0.5, 1.1) | 0.31 | 72 | −127 |
| Weibull model | 80 | 5 | 0.34 (0.3, 0.5) | 0.12 | 63 | −104 |
| Standard model (fixed) | 80 | 2 | 0.44 (0.3, 0.7) | 0.20 | 74 | **−140** |
| Weibull model (fixed) | 80 | 3 | 0.40 (0.3, 0.6) | 0.11 | 68 | −124 |
| **Gapyeong Stream** | | | | | | |
| Standard model | 100 | 4 | 1.5 (1.2, 1.7) | 0.73 | 139 | −260 |
| Weibull model | 100 | 5 | 1.5 (1.3, 1.7) | 0.73 | 135 | −246 |
| Standard model (fixed) | 100 | 2 | 1.4 (1.2, 1.7) | 0.64 | 135 | −261 |
| Weibull model (fixed) | 100 | 3 | 1.4 (1.2, 1.7) | 0.67 | 138 | **−262** |
| **Rob Roy Stream** | | | | | | |
| Standard model | 77 | 4 | 1.2 (1.0, 1.4) | 0.60 | 112 | −207 |
| Weibull model | 77 | 5 | 1.2 (1.0, 1.3) | 0.53 | 107 | −193 |
| Standard model (fixed) | 77 | 2 | 1.3 (1.0, 1.5) | 0.52 | 108 | −207 |
| Weibull model (fixed) | 77 | 3 | 1.2 (1.0, 1.4) | 0.55 | 111 | **−210** |

with an initial length distribution rather than an initial value problem with recently hatched nymphs from an initial length size class. In this case, $\alpha$ could be modified to allow for variable growth rate as nymphs mature. We anticipate that collection of nymph data from hatching would better constrain this modification to $\alpha$ in the parameter estimation routine.

This study parameterized the same model across several different streams, rather than characterizing a unique model tailored to each stream. There are two distinct advantages to this general—rather than specific—approach. First, we are able to compare similar parameters across different streams to facilitate the evaluation of factors that control mayfly length development [10, 19]. For example, a uniform increase of stream temperature across each site has the potential to affect the modeled length distribution by either forcing the temperature inside or outside the interval $T_L$ and $T_L + R$. As previously discussed, any shift in temperature can either skew or force the length distribution to remain stationary. Second, this model could be coupled to an ecological model forecast of synchronous emergence across a geographic area [18, 20, 21], which have the potential to inform environmental policy mitigating climate change effects [28]. Future model directions can focus on including additional sites or coupling this model to models of mayfly emergence across a geographic domain.

## 4    URLs

The data collected in this manuscript were digitized from data of mayfly length distributions from peer-reviewed scientific studies. The other data utilized was temperature data series for Rob Roy Stream was generated from data published on the web from the Norwegian Meteorological Institute, and interpolated to fit the study guidelines, as described in Table 1.

## 5    Author Contributions

Co-author Zobitz initially designed the study, evaluated model results, and contributed to the writing of the manuscript. Co-author Schilling contributed to the study design, the analysis of results, and the writing of the manuscript. All other authors collected data, contributed to the analysis of results and the writing of the manuscript.

## 6    Acknowledgements

This manuscript is the product of collaboration with two cohorts of undergraduate students in an upper division



Figure 6: Modeled and measured population distributions for the different sites considered, taken as an time ensemble. Results were generated by first randomly thinning the set of accepted parameter values by 50% for computational efficiency, next evaluating the mayfly model for each of the remaining sets of parameters. The thick line represents the median value, along with the 95% confidence interval. The red coloring is the Standard model (Equation 4) and the blue coloring is the Weibull model (Equation 5).

# References

[1] Benke, A. C., Huryn, A. D., Smock, L. A., & Wallace J. B. (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*, *18*(3), 308–343.

[2] Bottová, K., Derka, T., & Svitok, M. (2013). Population dynamics of mayflies in a constant temperature spring stream in West Carpathians. *Limnologica – Ecology and Management of Inland Waters*, *43*(6), 469–474.

[3] Braswell, B. H., Sacks, W. J., Linder, E., & Schimel, D. S. (2005). Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biology*, *11*(2), 335–355.

[4] Burnham, K. P., & Anderson, D. R., (Eds.) (2002). *Model Selection and Multimodel Inference.* New York, NY: Springer New York.

[5] Chadwick, M. A., & Feminella, J. W. (2001). Influence of salinity and temperature on the growth and production of a freshwater mayfly in the Lower Mobile River, Alabama. *Limnology and Oceanography*, *46*(3), 532–542.

[6] Clifford, H. F. (1982). Life cycles of mayflies (Ephemeroptera), with special reference to voltinism. *Quaestiones entomologicae*.

[7] Epele, L. B., Miserendino, M. L., & Pessacq, P. (2011). Life history, seasonal variation and production of Andesiops torrens (Lugo-Ortiz and McCafferty) and Andesiops peruvianus (Ulmer) (Ephemeroptera: Baetidae) in a headwater Patagonian stream. *Limnologica – Ecology and Management of Inland Waters*, *41*(1), 57–62.

[8] González, J. M., Basaguren, A., & Pozo, J. (2003). Life history, production and coexistence of two leptophlebiid mayflies in three sites along a northern Spain stream. *Archiv für Hydrobiologie*, *158*(3), 303–316.

[9] Harding, J. S., & Winterbourn, M. J. (1993). Life history and production of *Coloburiscus humeralis* (Ephemeroptera: Oligoneuriidae) in two South Island high-country streams, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *27*(4), 445–451.

[10] Harper, M. P., & Peckarsky, B. L. (2006). Emergence cues of a mayfly in a high-altitude stream ecosystem: Potential response to climate change. *Ecological Applications*, *16*(2), 612–621.

[11] Hurtt, G. C. & Armstrong, R. A. (1996). A pelagic ecosystem model calibrated with BATS data. *Deep-Sea Research Part II-Topical Studies in Oceanography*, *43*, 653–683.

[12] Huryn, A. D., & Wallace, J. B. (2000). Life history and production of stream insects. *Annual review of entomology*, *45*(1), 83–110.

[13] Huryn, D. (1996). Temperature-dependent growth and life cycle of Deleatidium (Ephemeroptera: Leptophlebiidae) in two high-country streams in New Zealand. *Freshwater Biology*, *36*(2), 351–361.

[14] Hwang, J. M., Yoon, T. J., Lee, S. J., & Bae, Y. J. (2009). Life history and secondary production of *Ephemera orientalis* (Ephemeroptera: Ephemeridae) from the Han River in Seoul, Korea. *Aquatic Insects*, *31*(sup1), 333–341.

[15] Johnson, A. C., Acreman, M. C., Dunbar, M. J., Feist, S. W., Giacomello, A. M., Gozlan, R. E., ... Williams, R. J. (2009). The British river of the future: How climate change and human activity might affect two contrasting river ecosystems in England. *Science of The Total Environment*, *407*(17), 4787–4798.

[16] Lee, S. J., Hwang, J. M., & Bae, Y. J. (2008). Life history of a lowland burrowing mayfly, Ephemera orientalis (Ephemeroptera: Ephemeridae), in a Korean stream. *Hydrobiologia*, *596*(1), 279–288.

[17] Logan, J. D., & Wolesensky, W. (2009). *Mathematical Methods in Biology* (1st edition). Hoboken, N.J: Wiley.

[18] Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., ... Schimel, D. S. (2011). Ecological forecasting and data assimilation in a data-rich era. *Ecological Applications*, *21*(5), 1429–1442.

[19] Nebeker, A. V. (1971). Effect of high winter water temperatures on adult emergence of aquatic insects. *Water Research*, *5*(9), 777–783.

[20] Newbold, J. D., Sweeney, B. W., & Vannote, R. L. (1994). A Model for Seasonal Synchrony in Stream Mayflies. *Journal of the North American Benthological Society*, *13*(1), 3–18.

[21] Niu, S., Luo, Y., Dietze, M. C., Keenan, T. F., Shi, Z., Li, J., & Chapman, F. S., III. (2014). The role of data assimilation in predictive ecology. *Ecosphere*, *5*(5), 1–16.

[22] R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

[23] RStudio Team (2015). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc.

[24] Schwartz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics*, *6*(2), 461–464.

[25] Sokal, R., & Rohlf, J. (1995). *Biometry*. W. H. Freeman & Company, New York.

[26] Sweeney, B. W., & Vannote, R. L. (1978). Size Variation and the Distribution of Hemimetabolous Aquatic Insects: Two Thermal Equilibrium Hypotheses. *Science*, *200*(4340), 444–446.

[27] van Vliet, M. T., Franssen, W. H., Yearsley, J. R., Ludwig, F., Haddeland, I., Lettenmaier, D. P., & Kabat, P. (2013). Global river discharge and water temperature under climate change. *Global Environmental Change*, *23*(2), 450–464.

[28] Wilby, R., Orr, H., Watts, G., Battarbee, R., Berry, P., Chadd, R., ... Extence, C. (2010). Evidence needed to manage freshwater ecosystems in a changing climate: Turning adaptation principles into practice. *Science of The Total Environment*, *408*(19), 4150–4164.

[29] Winterbourn, M. J., Cadbury, S., Ilg, C., & Milner, A. M. (2008). Mayfly production in a New Zealand glacial stream and the potential effect of climate change. *Hydrobiologia*, *603*(1), 211–219.

[30] Wright, L. L., Mattice, J. S., & Beauchamp, J. J. (1982). Effect of temperature and sex on growth patterns in nymphs of the mayfly Hexagenia hilineata in the laboratory. *Freshwater Biology*, *12*(6), 535–545.

[31] Zobitz, J., Desai, A., Moore, D., & Chadwick, M. (2011). A primer for data assimilation with ecological models using Markov Chain Monte Carlo (MCMC). *Oecologia*, *167*(3), 599–611.