Channel Width 2021

Thorley, J.L. & Irvine A.

Draft: 2021-04-10 14:43:36

The suggested citation for this analytic appendix is:

Thorley, J.L. & Irvine A. (2021) Channel Width 2021. A Poisson Consulting Analysis Appendix. URL: https://www.poissonconsulting.ca/f/1792764180.

Background

The primary goal of the current analyses is to answer the following question:

> How is stream channel width influenced by watershed area and other factors?

Data Preparation

The data were provided by New Graph Environment in the form an csv file and prepared for analysis using R version 4.0.4 (R Core Team 2020).

Key assumptions of the data preparation included:

- Data points with a channel width of 0 m, a channel width greater than 300 m, a watershed area less than 0.1 ha or a gradient less than 0 are unreliable and were excluded.

Statistical Analysis

Model parameters were estimated using Bayesian methods. The estimates were produced using JAGS (Plummer 2003) and STAN (Carpenter et al. 2017). For additional information on Bayesian estimation the reader is referred to McElreath (2016).

Unless stated otherwise, the Bayesian analyses used weakly informative normal and half-normal prior distributions (Gelman, Simpson, and Betancourt 2017). The posterior distributions were estimated from 1500 Markov Chain Monte Carlo (MCMC) samples thinned from the second halves of 3 chains (Kery and Schaub 2011, 38–40). Model convergence was confirmed by ensuring that the potential scale reduction factor $\hat{R} \leq 1.05$ (Kery and Schaub 2011, 40) and the effective sample size (Brooks et al. 2011) ESS ≥ 150 for each of the monitored parameters (Kery and Schaub 2011, 61).

The parameters are summarised in terms of the point *estimate*, *lower* and *upper* 95% credible limits (CLs) and 95% prediction limits (PLs) and the surprisal *s-value* (Greenland 2019). The estimate is the median (50th percentile) of the MCMC samples while the 95% CLs are the 2.5th and 97.5th percentiles. The 95% PLs are the 2.5th and 97.5th percentiles of individual channel widths based on the residual variation. The s-value can be considered a test of directionality. More specifically it indicates how surprising (in bits) it would be to discover that the true value of the parameter is in the opposite direction to the estimate. An s-value of 4.3 bits, which is equivalent to a p-value (Kery and Schaub 2011; Greenland and Poole 2013) of 0.05, indicates that the surprise would be equivalent to throwing 4.3 heads in a row. The condition that non-essential explanatory variables have s-values ≥ 4.3 bits provides a useful model selection heuristic (Kery and Schaub 2011).

Model adequacy was assessed via posterior predictive checks (Kery and Schaub 2011). More specifically, the number of zeros and the first four central moments (mean, variance, skewness and kurtosis) for the deviance residuals were compared to the expected values by simulating new residuals. In this context the s-value indicates how surprising each metric is given the estimated posterior probability distribution for the residual variation.

Where computationally practical, the sensitivity of the parameters to the choice of prior distributions was evaluated by increasing the standard deviations of all normal, half-normal and log-normal priors by an order of magnitude and then using $\hat{R}$ to test whether the samples where drawn from the same posterior distribution (Thorley and Andrusak 2017).

The results are displayed graphically by plotting the modeled relationships between particular variables and the response(s) with the remaining variables held constant. In general, continuous and discrete fixed variables are held constant at their mean and first level values, respectively, while random variables are held constant at their typical values (expected values of the underlying hyperdistributions) (Kery and Schaub 2011, 77–82). When informative the influence of particular variables is expressed in terms of the *effect size* (i.e., percent or n-fold change in the response variable) with 95% credible intervals (CIs, Bradford, Korman, and Higgins 2005).

The analyses were implemented using R version 4.0.4 (R Core Team 2020) and the `mbr` family of packages.

Model Descriptions

Channel Width The data were analysed using a power model. Key assumptions of the model include:

- The channel width varies with the upstream watershed area and mean annual precipitation.
- The channel width varies randomly by biogeoclimatic zone.
- The residual variation in channel width is log-normally distributed.

Model Templates

Channel Width

```
.data {
  int nObs;
  int nbgz;

  real width[nObs];
  real area[nObs];
  real precipitation[nObs];
  int bgz[nObs];
}
parameters {
  real b0;
  real bArea;
  real bPrecipitation;

  vector[nbgz] bbgz;
  real<lower=0> sbgz;

  real<lower=0> sWidth;
}
model {
  vector[nObs] eWidth;

  b0 ~ normal(0, 2);
  bArea ~ normal(0, 2);
  bPrecipitation ~ normal(0, 2);
  sbgz ~ normal(0, 2);
```

```
  bbgz ~ normal(0, sbgz);

  sWidth ~ normal(0, 2);

  for (i in 1:nObs) {
    eWidth[i] = exp(b0 + bArea * log(area[i])  + bPrecipitation * log(precipitation[i]) + bbgz[bgz[i]])
    width[i] ~ lognormal(log(eWidth[i]), sWidth);
  }
}
```

\end{lstlisting}

Block 1. Model description.

Results

Tables

Channel Width Table 1. Parameter descriptions.

Parameter	Description
`area[i]`	The upstream watershed area for the `i`th width (km^2)
`b0`	Intercept for `log(eWidth)`
`bArea`	Effect of `log(area)` on `b0`
`bbgz[i]`	Effect of `i`th biogeoclimatic zone on `b0`
`bgz[i]`	The biogeoclimatic zone for the `i`th width
`bPrecipitation`	Effect of `log(precipitation)` on `b0`
`eWidth[i]`	Expected value of `width[i]`
`precipitation[i]`	The mean annual precipitation for the `i`th width (m)
`sbgz`	SD of `bbgz`
`sWidth`	SD of residual variation in `width`
`width[i]`	The `i`th stream channel width (m)

Table 2. Model coefficients.

term	estimate	lower	upper	svalue
b0	-2.2383120	-2.4060742	-2.0546859	10.55171
bArea	0.3121556	0.3074218	0.3169608	10.55171
bPrecipitation	0.6546995	0.6322231	0.6775185	10.55171
sbgz	0.2194675	0.1359695	0.3898028	10.55171
sWidth	0.4907025	0.4863916	0.4952048	10.55171

Table 3. Model convergence.

n	K	nchains	niters	nthin	ess	rhat	converged
22990	5	3	500	1	172	1.033	TRUE

Table 4. Model posterior predictive checks.

moment	observed	median	lower	upper	svalue
zeros	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
mean	0.0000171	0.0002078	-0.0168776	0.0178864	0.0154610
variance	1.9990386	2.0010356	1.9650757	2.0369153	0.1391384
skewness	-0.3575806	-0.0001134	-0.0306088	0.0309393	10.5517083
kurtosis	1.3995289	-0.0007405	-0.0638227	0.0607513	10.5517083

Table 5. Model sensitivity.

n	K	nchains	niters	rhat_1	rhat_2	rhat_all	converged
22990	5	3	500	1.033	1.028	1.032	TRUE

Figures

Channel Width

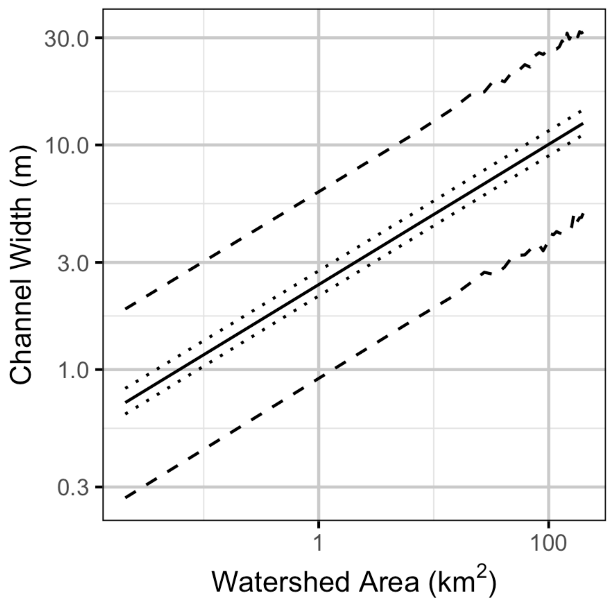

Figure 1. The predicted channel width by upstream water shed area on a log scale (with 95% CIs as dotted lines and 95% PI as dashed lines).

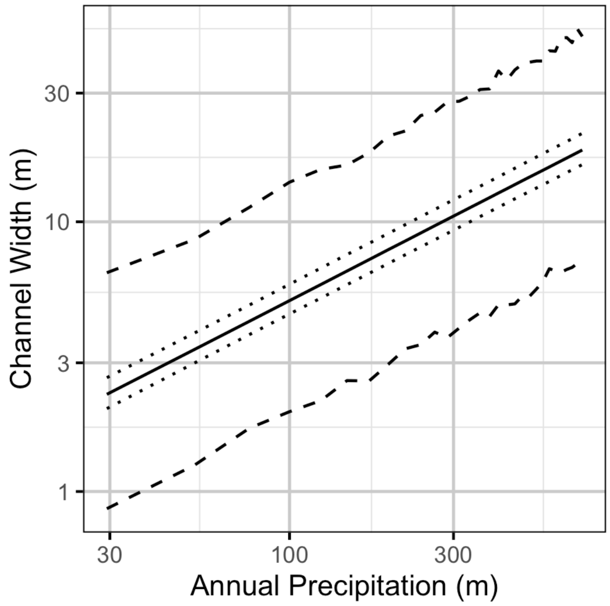

Figure 2. The predicted channel width by precipitation on a log scale (with 95% CIs as dotted lines and 95% PI as dashed lines).

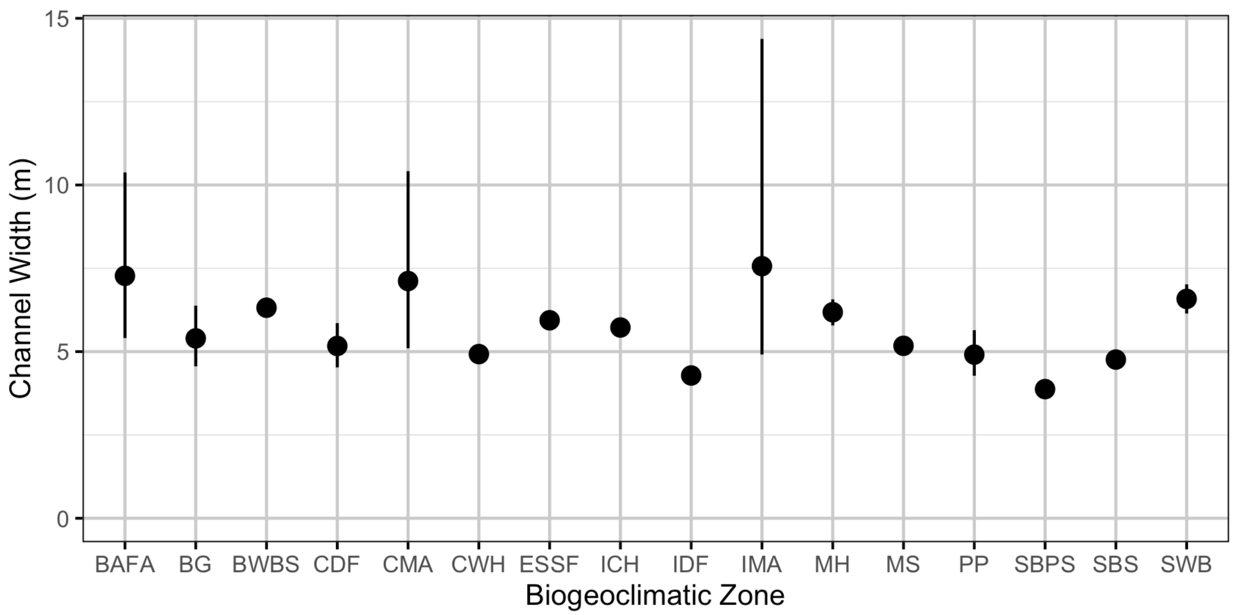

Figure 3. The predicted channel width by biogeoclimatic zone (with 95% CIs).

Acknowledgements

The organisations and individuals whose contributions have made this analytic appendix possible include:

- Hillcrest Geographics
  - Simon Norris

References

Bradford, Michael J, Josh Korman, and Paul S Higgins. 2005. “Using Confidence Intervals to Estimate the

Response of Salmon Populations (Oncorhynchus Spp.) To Experimental Habitat Alterations." *Canadian Journal of Fisheries and Aquatic Sciences* 62 (12): 2716–26. https://doi.org/10.1139/f05-179.

Brooks, Steve, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng, eds. 2011. *Handbook for Markov Chain Monte Carlo*. Boca Raton: Taylor & Francis.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "*Stan* : A Probabilistic Programming Language." *Journal of Statistical Software* 76 (1). https://doi.org/10.18637/jss.v076.i01.

Gelman, Andrew, Daniel Simpson, and Michael Betancourt. 2017. "The Prior Can Often Only Be Understood in the Context of the Likelihood." *Entropy* 19 (10). https://doi.org/10.3390/e19100555.

Greenland, Sander. 2019. "Valid p -Values Behave Exactly as They Should: Some Misleading Criticisms of p -Values and Their Resolution With s -Values." *The American Statistician* 73 (sup1): 106–14. https://doi.org/10.1080/00031305.2018.1529625.

Greenland, Sander, and Charles Poole. 2013. "Living with P Values: Resurrecting a Bayesian Perspective on Frequentist Statistics." *Epidemiology* 24 (1): 62–68. https://doi.org/10.1097/EDE.0b013e3182785741.

Kery, Marc, and Michael Schaub. 2011. *Bayesian Population Analysis Using WinBUGS : A Hierarchical Perspective*. Boston: Academic Press. http://www.vogelwarte.ch/bpa.html.

McElreath, Richard. 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman & Hall/CRC Texts in Statistical Science Series 122. Boca Raton: CRC Press/Taylor & Francis Group.

Plummer, Martyn. 2003. "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, edited by Kurt Hornik, Friedrich Leisch, and Achim Zeileis. Vienna, Austria.

R Core Team. 2020. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Thorley, Joseph L., and Greg F. Andrusak. 2017. "The Fishing and Natural Mortality of Large, Piscivorous Bull Trout and Rainbow Trout in Kootenay Lake, British Columbia (2008–2013)." *PeerJ* 5: e2874. https://doi.org/10.7717/peerj.2874.