Method Article

Hierarchical change-point regression models including random effects to estimate empirical critical loads for nitrogen using Bayesian Regression Models (brms) and JAGS

Tobias Roth a,b,1,*, Simon Tresch c,1, Enzai Du d, Sabine Braun c

a Zoological Institute, University of Basel, Vesalgasse 1, Basel CH-4051, Switzerland
b Hintermann & Weber, Austrasse 2a, Reinach CH-4153, Switzerland
c Institute for Applied Plant Biology, Benkenstrasse 254a, Witterswil CH-4108, Switzerland
d Faculty of Geographical Science, Beijing Normal University, Benkenstrasse 254a, Beijing 100875, China

ABSTRACT

The concept of critical loads is used in the framework of the Convention on Long-range Transboundary Air Pollution (UNECE) to define thresholds below which no damaging effects on habitats occur based on the latest scientific knowledge. Change-point regression models applied in a Bayesian framework are useful statistical tools to estimate critical empirical loads. While hierarchical study designs are common in ecological research, previous methods to estimate critical loads using change-point regression did not allow to analyse data collected under such a design. This method update provides an implementation of hierarchical data structure by including random effects such as study sites or as in this example tree species within the Bayesian approach of change-point regression models using two different approaches. The example data set is an European wide gradient study of the impact of climate change and air pollution on forest tree health assessed by foliar nutrient status of nitrogen (N) to phosphorus (P) from 10 different conifer tree species originated from 88 forest sites and 9 countries covering 22 years (1995-2017). Both modelling approaches using JAGS and Bayesian Regression Models using ‘Stan’ (brms) resulted in reasonable and similar estimations of the critical empirical load for nitrogen (CLempN) for temperate forests. These methodological examples of using different approaches of Bayesian change-point regression models dealing with random effects could prove useful to infer CLempN for other ecosystems and long-term data sets.

- Hierarchical change-point regression models are suitable for estimating critical empirical loads.
- The Bayesian framework of these models provides the inclusion of the current critical load and various confounding or modifying variables.
- Here we present two ways of implementing hierarchical data sets in Bayesian change-point regression models using JAGS and brms.

*Corresponding author.
E-mail address: roth@hintermannweber.ch (T. Roth).
1 These authors contributed equally to this paper.

https://doi.org/10.1016/j.mex.2022.101902
2215-0161/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
Specifications table

| Subject area                                   | Environmental Science |
|------------------------------------------------|-----------------------|
| More specific subject area                    | Forestry research     |
| Method name                                    | Revised Bayesian change-point regression models including random effects using brms and JAGS |
| Name and reference of original method         | Change-point models applied in a Bayesian context by [1] |
| Resource availability                         | Model data can be found in [2]. Codes including models and graphics are written in R [3] and given as an RMarkdown [4] output. The MCMC simulations of the change-point models were conducted using JAGS, version 4.3.0 [5], executed in R using rjags [6] and in STAN with brms [7]. |

Method details

Bayesian change-point regression model settings

The revised change-point models using either JAGS or brms are based on the current method of using Bayesian change-point models as in the example given by [8]. The Bayesian analysis is based on MCMC methods [9] with a similar setting used by [8] conducted using JAGS (version 4.3.0 [5] and executed in R using rjags [6]. Posterior distributions were based on parallel chains (t.n.chains = 2) with 100,000 iterations each (t.n.iter = 100,000), discarding the first 50,000 values (t.n.burnin = 50,000) and thinning the remainder by 2 (t.n.thin = 2). The two parallel chains were used to assess convergence with Gelman and Rubin’s diagnostics [10] using trace and marginal density plots (Fig. 2) with the R function plot(coda) and scale reduction factors with the R function gelman.diag(coda) using the R package coda [11]. We are presenting the results as the mean (point estimate) and the 2.5% and 97.5% quantiles (95% credible intervals (CI), [12]) of the posterior distribution following [13].

Model data set

The used data set is a European wide gradient study of the impact of climate change and air pollution (N deposition) on forest tree health assessed by foliar nutrient status N:P of different different conifer species (n=10) from the years 1995-2017 [2]. N deposition are based on the EMEP MSC-W model with a 0.1° spatial resolution [14]. Mean annual temperatures (MAT) are based on the E-OBS climate data set with a 0.25° spatial resolution [2].

Prior settings

We used the approved critical load for coniferous woodland, which is currently set at 5-15 kg N ha⁻¹ a⁻¹ [15], to construct an informative prior (CL_emp,N) assuming a normal distribution with the approved critical load as its mean and half the range as its standard deviation (Normal(mean = 10, sd = 5)) and vague priors with mean = 0 and sd = 2 for fixed effects such as the mean annual temperature (MAT) in this example.
Bayesian change-point regression model using JAGS (BCR_JAGS)

According to the current Bayesian framework of change-point regression models by [1], we used a linear mixed effect model (c.f. data distribution in Fig. 1) with identity-link function. We described the variation of measured foliar N:P ratio \( (NP_i) \) between \( i = 1, ..., N \) sampling sites using a Normal distribution with expected foliar N:P ratio \( \lambda_i \):

\[
NP_i \sim \text{Normal}(\lambda_i)
\]

and the expected foliar N:P ratio \( \lambda_i \) as identity link function expressed as:

\[
\lambda_i = \beta_0 + \sum_{k=1}^{K} \beta_k X_{i,k} + \alpha_i
\]

with \( \beta_0 \) as intercept and \( \beta_k \) as linear slopes for \( k = 1, ..., K \) confounding variables at sampling site \( i \) with covariate value \( X_{i,k} \) [16]. According to [1] we added \( \alpha_i \) as the effect of N deposition \( N_i \) on the expected foliar N:P ratio \( \lambda_i \):

\[
\alpha_i = \begin{cases} 
0 & \text{if } N_i < \text{CL} \\
\beta_N (N_i - \text{CL}) & \text{if } N_i \geq \text{CL}
\end{cases}
\]

assuming no effect of N deposition if N deposition \( N_i \) is below the critical load (CL) and a linear change (\( \beta_N \)) of \( \lambda_i \) with increasing N deposition if \( N_i \) is equal or above the critical load (CL).

The random effect term (lines 20-22 in the JAGS model: JAGS_change_point_model_random_effect.R) was defined as:

\[
\lambda_i = \beta_{0.5} + \sum_{k=1}^{K} \beta_k X_{i,k} + \alpha_i
\]

\( \beta_{0.5} \sim \text{norm}(\mu, \sigma) \) with the intercept \( \beta_{0.5} \) that varies between conifer tree species in this example. The differences between the tree species is modelled using a normal distribution, which is the definition of a random effect.

BCR_JAGS model diagnostics

The convergence diagnostics using Gelman and Rubin’s [10] trace and marginal density plots (Fig. 2) showed a good convergence of the two chains. Also the smoothed density plots showed rather balanced histograms of the trace plot values for this model.

Estimated change-point using JAGS

The effect plot of the estimated change-point including the random effect of conifer tree species is shown in Fig. 3. The change-point, using the underlying data on foliar N:P ratio and the Bayesian change-point regression model BCR_JAGS, is 5.7 ± 4.8 (Tab. 1).
Bayesian change-point regression model using brms (BCR_brms)

With the package “brms” [7] we fitted a Bayesian non-linear multivariate multilevel model with random intercept ($\beta_0$) and different $\text{CL}_\text{empN}$ between the tree species:

```
brmsformula(  
  dependvar ~ $\beta_0 + \beta_1 * \text{fixeff}$ +  
  step(Ndep - $\text{CL}_\text{empN}$) * betaN * (Ndep - $\text{CL}_\text{empN}$),  
  $\text{CL}_\text{empN}$ 1 + (1 | random),  
  beta0 1 + (1 | random),  
  beta1 + betaN ~ 1,  
  nl = TRUE, family = gaussian)
```

With the same priors and model settings as before in the model BCR_JAGS (chains = 2, iter = 100000, thin = 2, warmup = 50000, cores = 4).
Fig. 3. Change-point regression (model BCR_JAGS) of foliar N:P ratios including the fixed effect MAT and tree species as random effect. The grey line represents the estimated critical load (Table 1) of N including the corresponding 95% credible intervals as dotted lines. Points are measurements from [2] and are coloured according to conifer tree species. The black line is the estimated change-point regression including 95% credible intervals.

|            | Estimated $Cl_{emp}$ N (kg ha$^{-1}$ a$^{-1}$) | Mean | SD   | CI 2.5% | CI 97.5% |
|------------|-----------------------------------------------|------|------|---------|----------|
| BCR_JAGS   | 5.71                                          | 4.81 | -4.69| 14.81   |          |
| BCR_brms   | 9.06                                          | 4.31 | 0.49 | 17.76   |          |

**BCR_brms model diagnostics**

The MCMC diagnostics showed a good convergence of the two chains. The posterior distributions are centred around one peak value (Fig. 4).

**Estimated change-point using brms**

The estimated change-point using the model BCR_JAGS is 9.0±4.3 (Table 1) with an estimated Bayesian $R^2 = 0.42$ [17]. The effect plot of the change-point is shown in Fig. 5.
Fig. 4. Bayesian change-point regression model convergence diagnostics of the model BCR_brms. Right: Trace plots showing the two chains. Left: Density plot of the posterior distributions.

**Method validation**

The two presented ways of modelling change-point regressions including random effects, with JAGS (BCR_JAGS) and with the brm function from the brms [7] package (BCR_brms), resulted in similar estimated $\text{CL}_{\text{emp}}N$ (Table 1), latter being slightly higher with the model (BCR_brms). The estimated CL from the model (BCR_brms) including the non-linearity and the random intercepts of $\beta_0$ made the outcome more realistic compared to the underlying data (Fig. 5).

In addition, a simple boxplot comparison (pinpoint method applied by [15]) of different groups of N deposition is suggesting a change in foliar N:P ratio between the first group of 0-5 kg ha$^{-1}$ a$^{-1}$ and the second group of 5-10 kg ha$^{-1}$ a$^{-1}$ (Fig. 6), which verifies the estimated $\text{CL}_{\text{emp}}N$ of both models.
Fig. 5. Change-point regression (model BCR_brms) of foliar N:P ratios including the confounding factor mean annual temperature (MAT) and tree species as random effect. The grey line represents the estimated critical load (Tab. 1) of N including the corresponding 95% credible intervals as dotted lines. Points are measurements from [2] and are coloured according to conifer tree species. The black line is the estimated change-point regression including 95% credible intervals.

Additional information

Introduction to critical empirical loads and change-point regression models

The deposition of atmospheric nitrogen (N) is a major threat to biodiversity and ecosystem functioning globally [18–22]. The accumulation of N over time is one of the main drivers of changes in species distribution across many habitats, especially for habitats with low management regimes such as natural forests [15]. The impacts of N deposition are known for certain terrestrial habitats, but many still remains uncertain [15]. Critical loads of N has been developed in the framework of the Convention on Long-range Transboundary Air Pollution [15,23]. These critical loads are defined as thresholds below which damaging effects on specific habitats do not occur based on the latest scientific knowledge [24]. Critical loads of N are often defined based on negative effects on plant diversity [20], as they are often lower compared to critical loads of N from soil processes such as soil acidification or N leaching [25]. Empirical critical loads can be determined with the use of N addition or reduction experiments or in gradient studies covering a gradient of N deposition. Latter is a more holistic approach taking into account the variability of natural habitats. However, gradient studies can also have various limitations for instance they should use a sufficient spatial scale of the air pollutant assessed and they should take into account the most important modifying factors [26]. In addition, an appropriate framework of regression model should be selected, suitable for the selected study design. Bayesian change-point regression models have been shown to overcome some difficulties of large spatial variation found in natural habitats by taking into account confounding factors such as climatic
Fig. 6. Changes in foliar N:P ratio with increasing N deposition values. The group-wise boxplot comparison (pinpoint method according to [15]) highlights a suggested change in foliar N:P ratio between the first (0-5 kg N ha\(^{-1}\) a\(^{-1}\)) and the second group (5-10 kg N ha\(^{-1}\) a\(^{-1}\)).

variation or different soil types [23,27], allowing a more accurate estimation of the critical load [28]. Therefore, change-point regression models applied in a Bayesian framework are useful statistical tools in estimating critical empirical loads [1].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper is a collaborative piece of work between ST and SB from the Institute for Applied Plant Biology providing the idea for the adaption of the current change-point method in the frame work of the revision of the critical empirical load for nitrogen and TR from Hintermann & Weber AG providing the statistical back ground. ED provided the data and edited the manuscript. This project was supported by the Swiss Federal Office for the Environment (FOEN).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2022.101902
References

[1] T. Roth, L. Kohli, B. Rihm, R. Meier, B. Achermann, Using change-point models to estimate empirical critical loads for nitrogen in mountain ecosystems, Environ. Pollut. 220 (2017) 1480–1487, doi:10.1016/j.envpol.2016.10.083.

[2] E. Du, M. van Doorn, W. v. Dries, Spatially divergent trends of nitrogen versus phosphorus limitation across European forests, Sci. Total Environ. 771 (2021) 145391, doi:10.1016/j.scitotenv.2021.145391.

[3] R Core Team, R Development Core Team: R: A Language and Environment for Statistical Computing, 2021, http://www.r-project.org/.

[4] J.J. Allaire, Y. Xie, J. McPherson, J. Luraschi, K. Ushey, A. Atkins, H. Wickham, J. Cheng, W. Chang, R. Iannone, rmarkdown: Dynamic Documents for R, 2021. https://github.com/rstudio/rmarkdown.

[5] M. Plummer, JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling, 3rd Int. Work. Distrib. Stat. Comput. (DSC 2003); Vienna, Austria 124 (2003).

[6] M. Plummer, rjags: Bayesian Graphical Models using MCMC. https://cran.r-project.org/package=rjags.

[7] P.-C. Bürkner, brms : An R Package for Bayesian Multilevel Models Using Stan, J. Stat. Softw. 80 (1) (2017), doi:10.18637/jss.v080.i01.

[8] T. Roth, P. Sprau, M. Naguib, V. Amrhein, Sexually selected signaling in birds: A case for Bayesian change-point analysis of behavioral routines, Auk 129 (4) (2012) 660–669, doi:10.1525/auk.2012.12041.

[9] W.A. Link, E. Cam, J.D. Nichols, E.G. Cooch, Of Bugs and Birds: Markov Chain Monte Carlo for Hierarchical Modeling in Wildlife Research, J. Wildl. Manage. 66 (2) (2002) 277, doi:10.2307/3803160.

[10] S.P. Brooks, A. Gelman, General Methods for Monitoring Convergence of Iterative Simulations, J. Comput. Graph. Stat. 7 (4) (1998) 434–455, doi:10.1080/10618600.1998.10474787.

[11] M. Plummer, N. Best, K. Cowles, K. Vines, CODA: Convergence Diagnosis and Output Analysis for MCMC, R News 6 (1) (2006) 7–11, https://journal.r-project.org/archive/.

[12] A. Gelman, S. Greenland, Are confidence intervals better termed uncertainty intervals? BMJ 366 (2019) i5381, doi:10.1136/bmj.i5381.

[13] F. Korner-Nievergelt, T. Roth, S. Von Fellen, J. Guelat, B. Almasi, P. Korner-Nievergelt, Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and Stan, Elsevier, 2015, doi:10.1016/C2013-0-23227-X.

[14] H. Fagerl, S. Tsyro, J.E. Jonson, A. Nyiri, D. Simpson, P. Wind, A. Benedictow, H. Klein, Q. Mu, B.R. Denby, E.G. Waersted, EPME - Status Report 1/2020 - Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components, Technical Report, Norwegian Meteorological Institute, 2020.

[15] R. Bobbink, S. Braun, A. Nordin, S. Power, K. Schütz, J. Strengbom, M. Weijters, H. Tomasson, in: Review and revision of empirical critical loads and dose-response relationships. Proceedings of an expert workshop, Noordwijkhout, 23-25 June 2010, CCE, 2011. http://www.ivrim.nl/bibliotheek/rapporten/680359002.pdf

[16] A. Gelman, J. Hill, Data analysis using regression and multilevel/hierarchical models, Cambridge university press, 2006.

[17] A. Gelman, B. Goodrich, J. Gabry, A. Vehtari, R-squared for Bayesian Regression Models, Am. Stat. 73 (3) (2019) 307–309, doi:10.1080/00031305.2018.1549100.

[18] O.E. Sala, F. Stuart Chapin, III, J.J. Armesto, E. Berlow, J. Bloomfield, R. Dirzo, E. Huber-Sanwald, L.F. Hueneke, R.B. Jackson, A. Kinzig, R. Leemans, D.M. Lodge, H.A. Mooney, M. Oesterheld, N.L. Poff, M.T. Sykes, B.H. Walker, M. Walker, D.H. Wall, Global Biodiversity Scenarios for the Year 2100, Science (80-.) 287 (5459) (2000) 1770–1774, doi:10.1126/science.287.5459.1770.

[19] C.J. Stevens, How long do ecosystems take to recover from atmospheric nitrogen deposition? Biol. Conserv. 200 (2016) 160–167, doi:10.1016/j.biocon.2016.06.005.

[20] R. Bobbink, K. Hicks, J. Galloway, T. Sprunger, R. Alkemade, M. Ashmore, M. Bustamante, S. Cinderby, E. Davidson, F. Dentener, B. Emmett, J.-W. Erismann, M. Fenn, F. Gilliam, A. Nordin, L. Pardo, W. De Vries, Global assessment of nitrogen deposition effects on terrestrial plant diversity: a synthesis, Ecol. Appl. 20 (1) (2010) 30–59, doi:10.1890/10-1140.1.

[21] C.J. Stevens, T.I. David, J. Storkey, Atmospheric nitrogen deposition in terrestrial systems: Its impact on plant communities and consequences across trophic levels, Funct. Ecol. 32 (7) (2018) 1757–1769, doi:10.1111/1365-2435.13063.

[22] IPBES, Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Technical Report, 2019, doi:10.5281/zenodo.3831673. Bonn, Germany

[23] W. de Vries, M. Posch, H.U. Sverdrup, T. Larssen, H.A. de Wit, R. Bobbink, J.-P. Hettelingh, Geochemical Indicators for Use in the Computation of Critical Loads and Dynamic Risk Assessments, in: Crit. Loads Dyn. Risk Assessments, 2015, pp. 15–58, doi:10.10789-94-017-9508-1_2.

[24] J. Nilsson, P. Grennfelt, N.C. of Ministers, Critical loads for sulphur and nitrogen. Report from a workshop held at Skokloster, Sweden, 19-24 March, 1988, 1988.

[25] E. Du, Effects of Nitrogen Deposition on Forest Ecosystems, in: Handb. Air Qual. Clim. Chang., Springer Singapore, Singapore, 2022, pp. 1–23, doi:10.1007/978-988-18-15227-8_27-1.

[26] S. Braun, C. Schindler, B. Rihm, Growth trends of beech and Norway spruce in Switzerland: The role of nitrogen deposition, ozone, mineral nutrition and climate, Sci. Total Environ. 599-600 (2017) 637–646, doi:10.1016/j.scitotenv.2017.04.230.

[27] P.M. Goffman, J.S. Baron, T. Blett, A.J. Gold, I. Goodman, L.H. Gunderson, B.M. Levinson, M.A. Palmer, H.W. Paerl, G.D. Peterson, N.L. Poff, D.W. Rejeski, J.F. Reynolds, M.G. Turner, K.C. Weathers, J. Wiens, Ecological Thresholds: The Key to Successful Environmental Management or an Important Concept with No Practical Application? Ecosystems 9 (1) (2006) 1–13, doi:10.1007/s10021-003-0142-z.

[28] B. Beckage, L. Joseph, P. Belisle, D.B. Wolfson, W.J. Platt, Bayesian change-point analyses in ecology. New Phytol. 174 (2) (2007) 456–467, doi:10.1111/j.1469-8137.2007.01991.x.