

Ecological Threshold and Environmental Management: a note on statistical methods for detecting thresholds

Appendix

Song S. Qian^{* †}

November 1, 2013

This appendix includes (1) statistical details of the three linear threshold models and computational methods, (2) plots not included in the paper, and (3) residual normal quantile-quantile (Q-Q) plots.

1 Change Point Models

A change point model can be expressed as follows.

$$y_i \sim \begin{cases} \pi(y_i|\theta_1, x_i) & \text{if } x_i < \phi \\ \pi(y_i|\theta_2, x_i) & \text{if } x_i \geq \phi \end{cases} \quad (1)$$

^{*}Corresponding author, Phone: 419.530.4230, Fax: 419.530.4421, email Song.Qian@utoledo.edu

[†]Department of Environmental Sciences, University of Toledo, Toledo, OH 43606-3390, USA.

where y_i is the i th observed response variable value and x_i is the corresponding environmental variable (predictor) value, π represents a probability distribution. Equation (1) indicates that the distribution of y is parameterized by θ and/or x . The parameter values change (from θ_1 to θ_2) when the environmental variable crosses a change point ϕ . The family of linear change point model with one predictor variable include the step function (SF):

$$y_i \sim \begin{cases} N(\mu_1, \sigma_1^2) & \text{if } x_i < \phi \\ N(\mu_2, \sigma_2^2) & \text{if } x_i \geq \phi, \end{cases} \quad (2)$$

the broken stick model (BS):

$$y_i \sim \begin{cases} N(\beta_1 + \beta_2(x_i - \phi), \sigma^2) & \text{if } x_i < \phi \\ N(\beta_1 + (\beta_2 + \delta)(x_i - \phi), \sigma^2) & \text{if } x_i \geq \phi, \end{cases} \quad (3)$$

and the disjointed broken stick model (dBS):

$$y_i \sim \begin{cases} N(\beta_1 + \beta_2(x_i - \phi), \sigma^2) & \text{if } x_i < \phi \\ N((\beta_1 + \delta_1) + (\beta_2 + \delta_2)(x_i - \phi), \sigma^2) & \text{if } x_i \geq \phi, \end{cases} \quad (4)$$

where y is the response variable, x is the predictor (environmental gradient).

As in traditional linear models, normality is assumed. The normality assumption can be assessed graphically using a normal Q-Q plot of the model's residuals.

2 Computation

I used Markov chain Monte Carlo (MCMC) simulation for parameter estimation. All parameters were assigned flat priors. MCMC were implemented using JAGS. The JAGS codes for the three change point models are listed below.

```

## the dBS model

model{

  for (i in 1:n){

    y[i] ~ dnorm(y.hat[i], prec)

    y.hat[i] <- (beta[1]+delta[1]*step(x[i]-phi)) +

                (beta[2]+delta[2]*step(x[i]-phi))*(x[i]-phi)

  }

  for (i in 1:2){

    delta[i] ~ dnorm(0, 0.01)

    beta[i] ~ dnorm(0, 0.001)

  }

  prec <- pow(sigma, -2)

  sigma ~ dunif(0, 10)

  phi ~ dunif(1,u)

}

```

```

## The BS model

model{

  for (i in 1:n){

    y[i] ~ dnorm(y.hat[i], prec)

    y.hat[i] <- beta[1] +

                (beta[2]+delta*step(x[i]-phi))*(x[i]-phi)

  }

  for (i in 1:2){

```

```

    beta[i] ~ dnorm(0, 0.0001)
  }

  prec <- pow(sigma, -2)

  sigma ~ dunif(0, 10)

  delta ~ dnorm(0, 0.0001)

  phi ~ dunif(1,u)
}

## the SF model

model{

  for (i in 1:n){

    y[i] ~ dnorm(y.hat[i], prec[J[i]])

    y.hat[i] <- beta +

      delta*step(x[i]-phi)

    J[i] <- 1+step(x[i]-phi)
  }

  delta ~ dnorm(0, 0.0001)

  beta ~ dnorm(0, 0.0001)

  phi ~ dunif(1,u)

  for (j in 1:2){

    prec[j] <- pow(sigma[j], -2)

    sigma[j] ~ dunif(0, 50)
  }
}

```

3 Additional Plots

Plots are grouped by response variables. I used these plots to learn whether the data show evidence supporting the step function model. Because a statistical model is optimally fit to the data, just observing the usual diagnostic plots or statistics may not be sufficient. I used the multiple alternative models approach and judged whether there is evidence supporting one particular model by examining three pieces of information.

- The residual plot is commonly used to check for assumptions about a model's residual distribution. In this case, we assume that residuals should have a normal distribution with mean 0 and a constant variance. A residual plot is also effective in detecting patterns that can be used for suggesting further model modification. When comparing residual plot of the step function model to the same from alternative models, we should look for patterns in two areas divided by the estimated change point (if any). When the underlying pattern is monotonic, a step function model will almost always find a change point and the resulting residuals plot will likely show two areas of increasing or decreasing trend.
- Superposing the fitted model to the data scatter plot will allow us to compare the fitted and the observed. This plot is usually not very informative when only a single model is examined. When comparing multiple alternative models, we are more likely to distinguish a more suitable model from a less suitable one. In many cases, all alternative models seem to fit the same data equally well, which may suggest that the appropriate model is not included in the list of alternatives.
- For threshold models, I also show the posterior distributions of the estimated change point. A well-defined (concentrated) change point distribution is often

used as an indication of the existence of a threshold. The dBS model represents the generalization of the HS and SF models. If the SF model is the appropriate model, the estimated change point distributions for the dBS model and the SF model should be similar. Likewise, if the HS model is appropriate, the change point distributions of the dBS model and the HS model should be similar.

In short, the process of model assessment should be a process of carefully examining all available information.

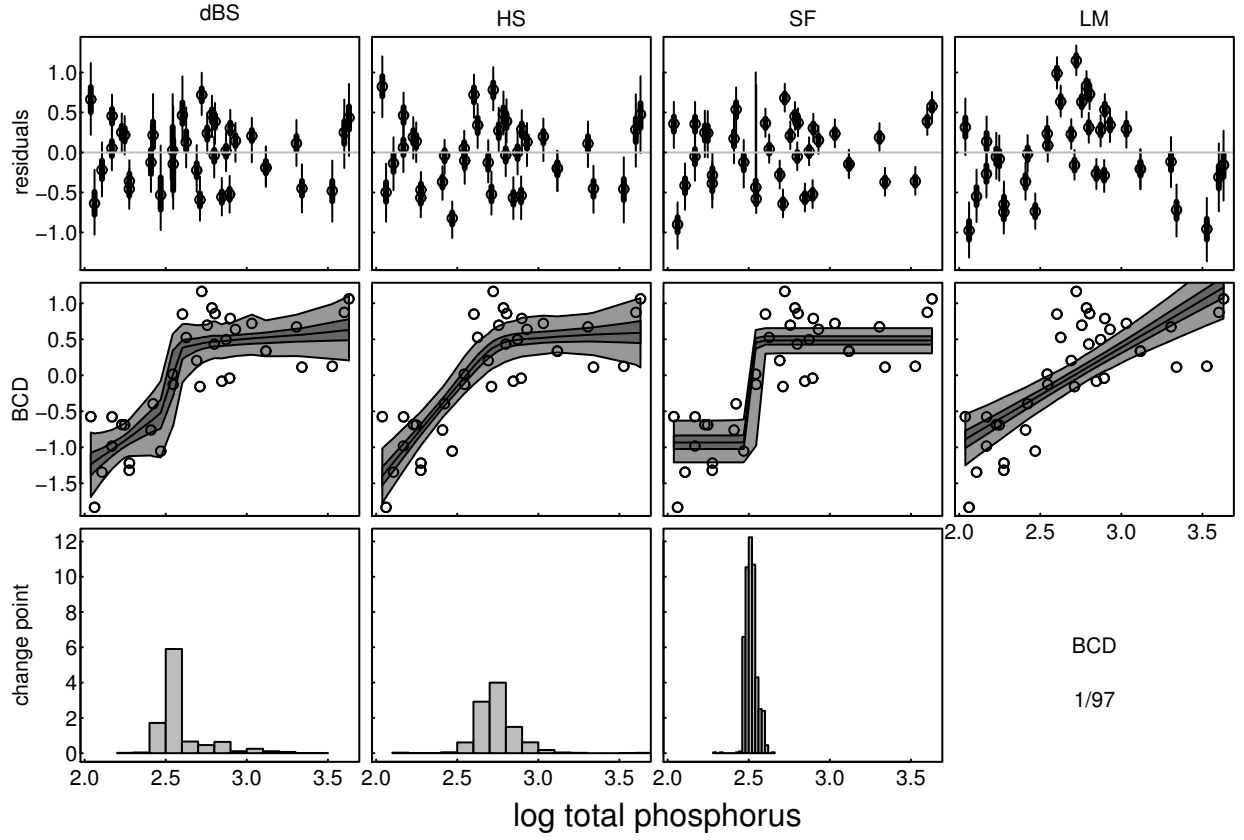


Figure 1: BCD data (collected in January 1997) show that the SF is likely the most suitable model. The residual plots (top row) show that the linear model is inadequate as a nonlinear pattern is clearly discernable. The fitted and observed plots (middle row) show that all four models fit to the data well, especially the SF model. The estimated change point posterior distributions (bottom row) show that all three change point models have a well-defined change point. But the SF model change point peak is much closer to the dBS model peak than the same of the HS model. A sigmoidal dose-response model is, however, also likely.

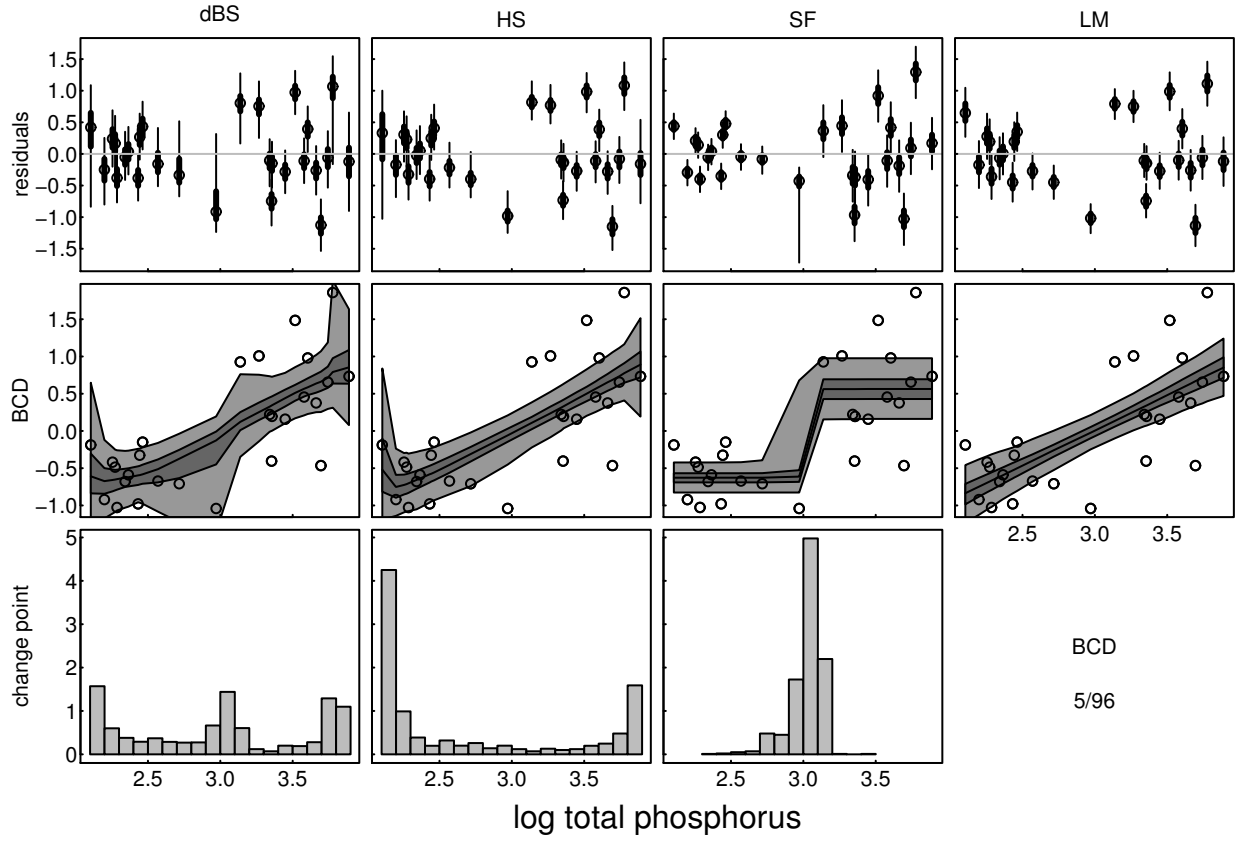


Figure 2: BCD data (May 1996) do not show support for the SF model. The data had a gap in the middle of the TP gradient resulting in two clusters. As a result, both the SF model and the linear model seem to fit the data well. The SF model resulted in a concentrated change point posterior distribution, but the dBS model did not.

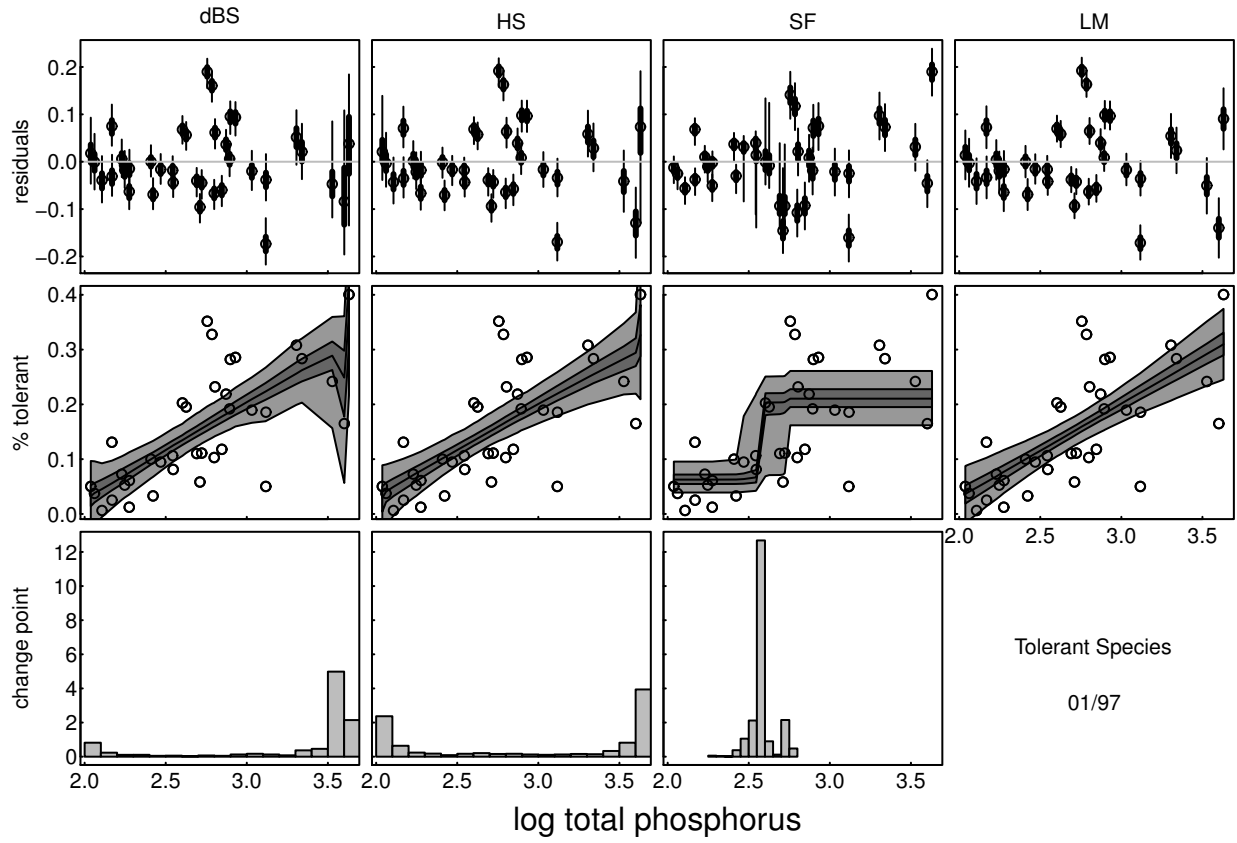


Figure 3: % Tolerant individual data (January 1997) show a strong support for the linear model as both the fitted HS and dBS models reduced to the linear model. The residual plots show that residual variance increases as TP increases. The concentrated change point distribution from the SF model is likely an artifact.

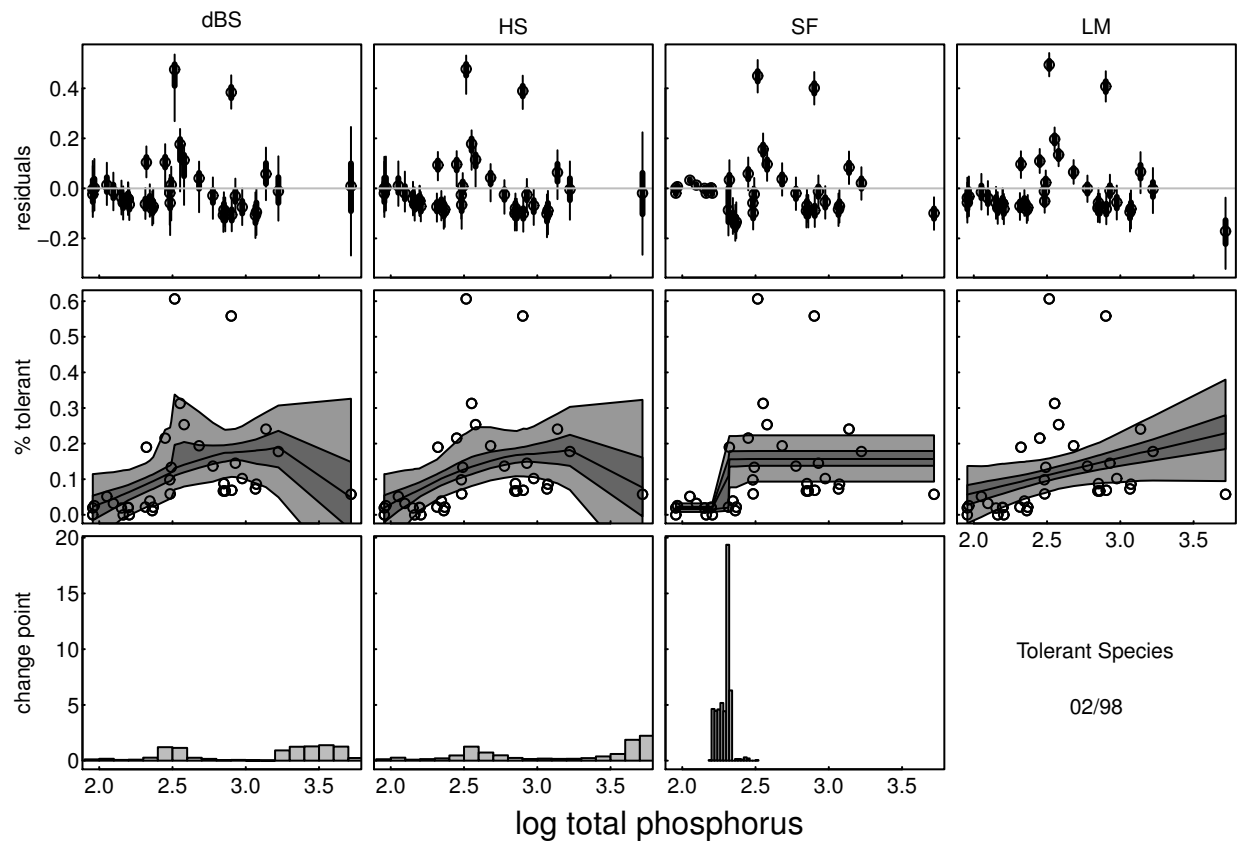


Figure 4: % Tolerant individual data (February 1998) show a strong non-linear response pattern. All four alternatives are unlikely to be appropriate. However, the SF model resulted in a concentrated change point posterior distribution.

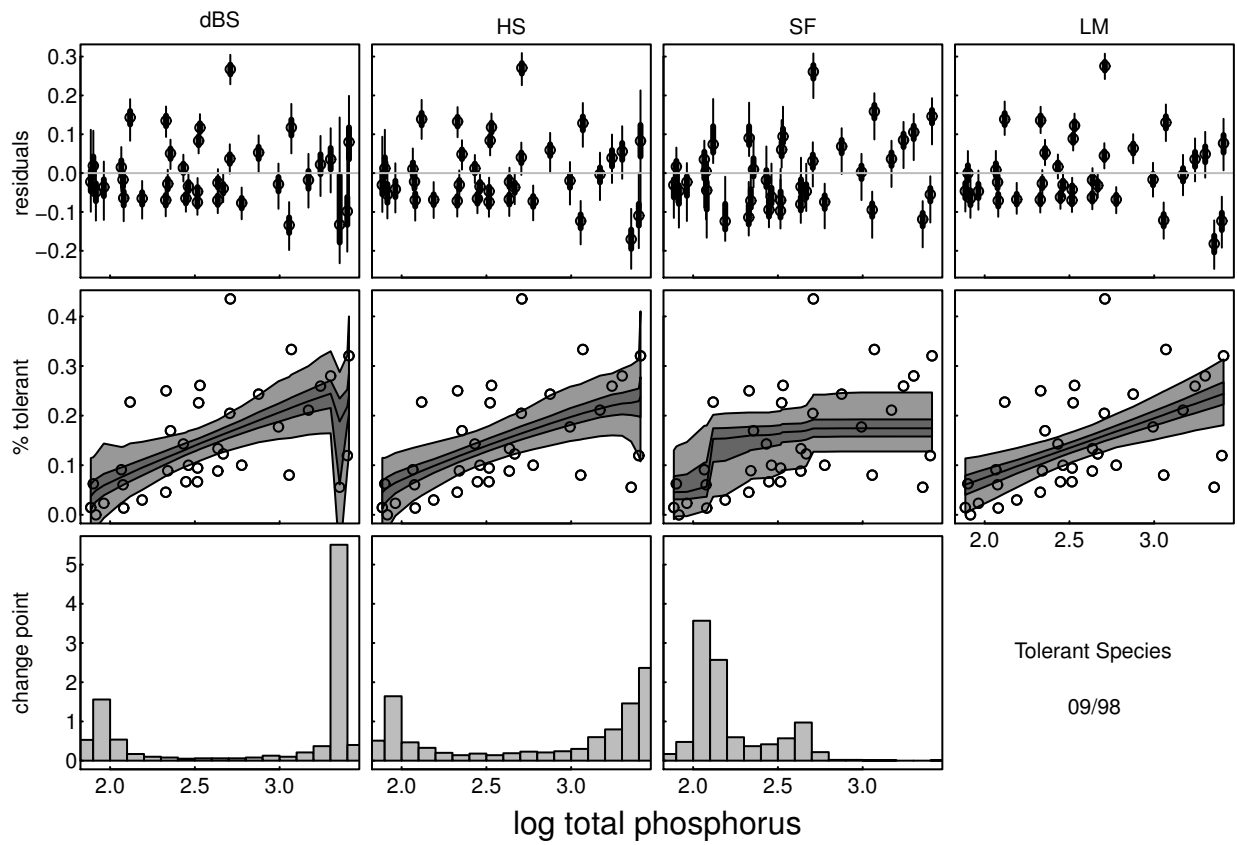


Figure 5: % Tolerant individual data (September 1998) do not show support for the SF model. Figures 3 – 5 show an increasing trend in residual variance as TP increases. The next four plots show the logit transformed % tolerant individual data, a transformation that often results in a better concordance to the normality assumption.

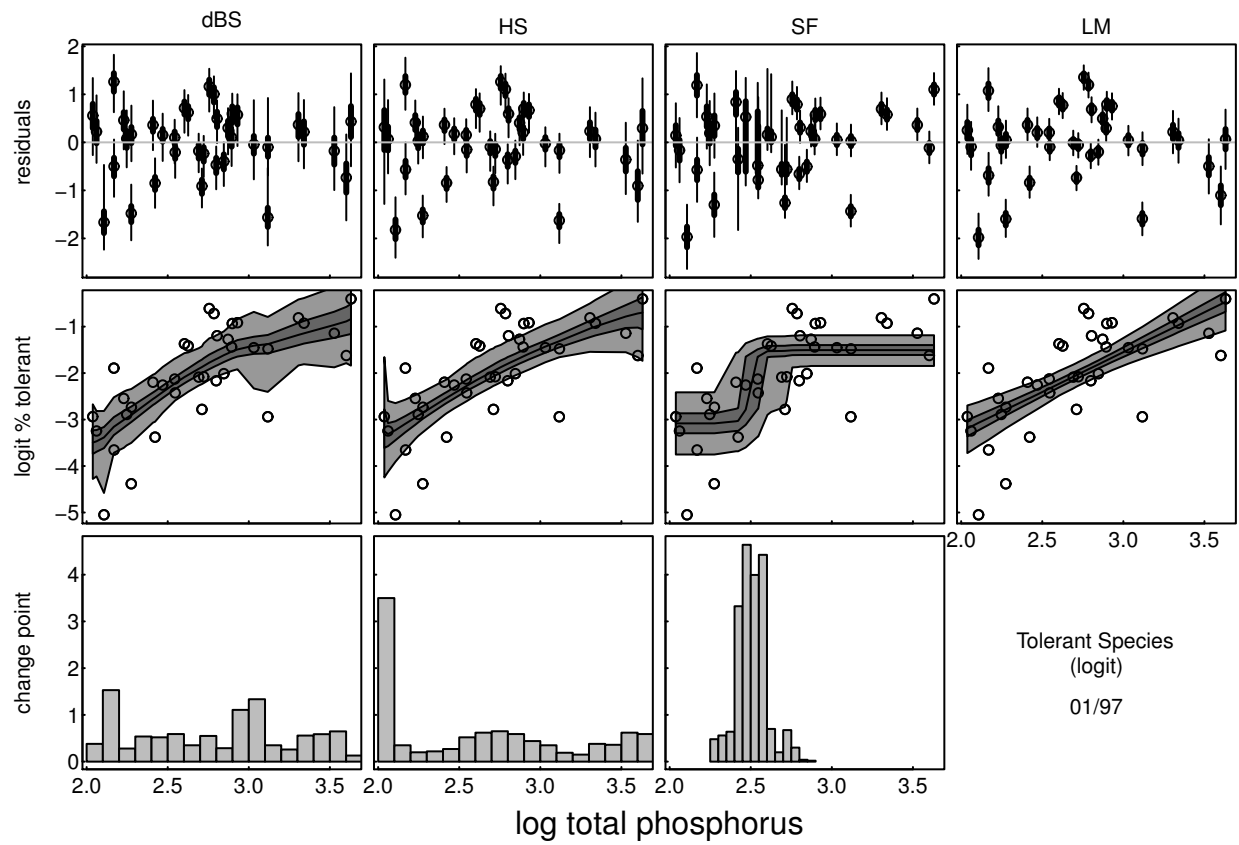


Figure 6: Logit transformed % tolerant individual data (January 1997) have a stable residual variance and a nonlinear model seems to be more appropriate than the four alternative models

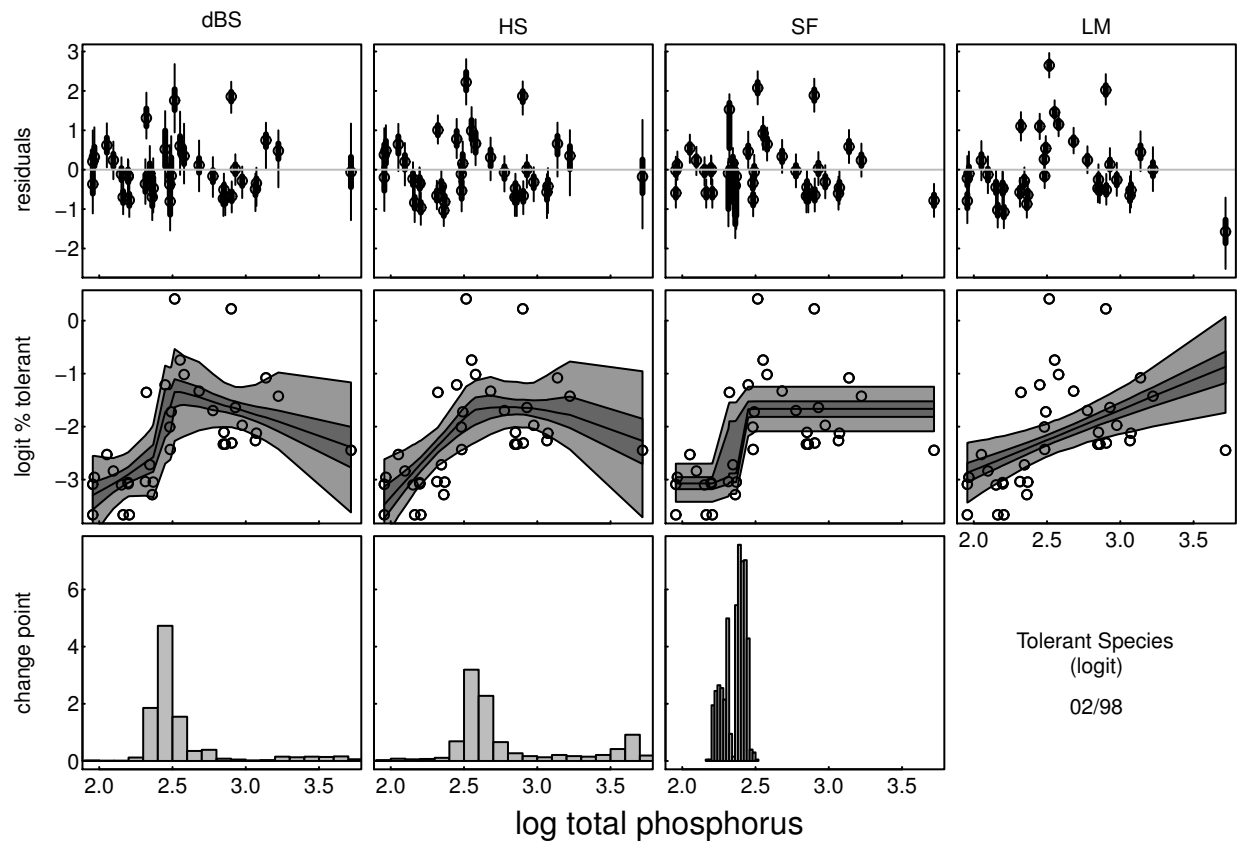


Figure 7: Logit transformed % tolerant individual data (February 1998) show a strong nonlinear (likely quadratic) relationship. It is difficult to pick the SF model over the HS model.

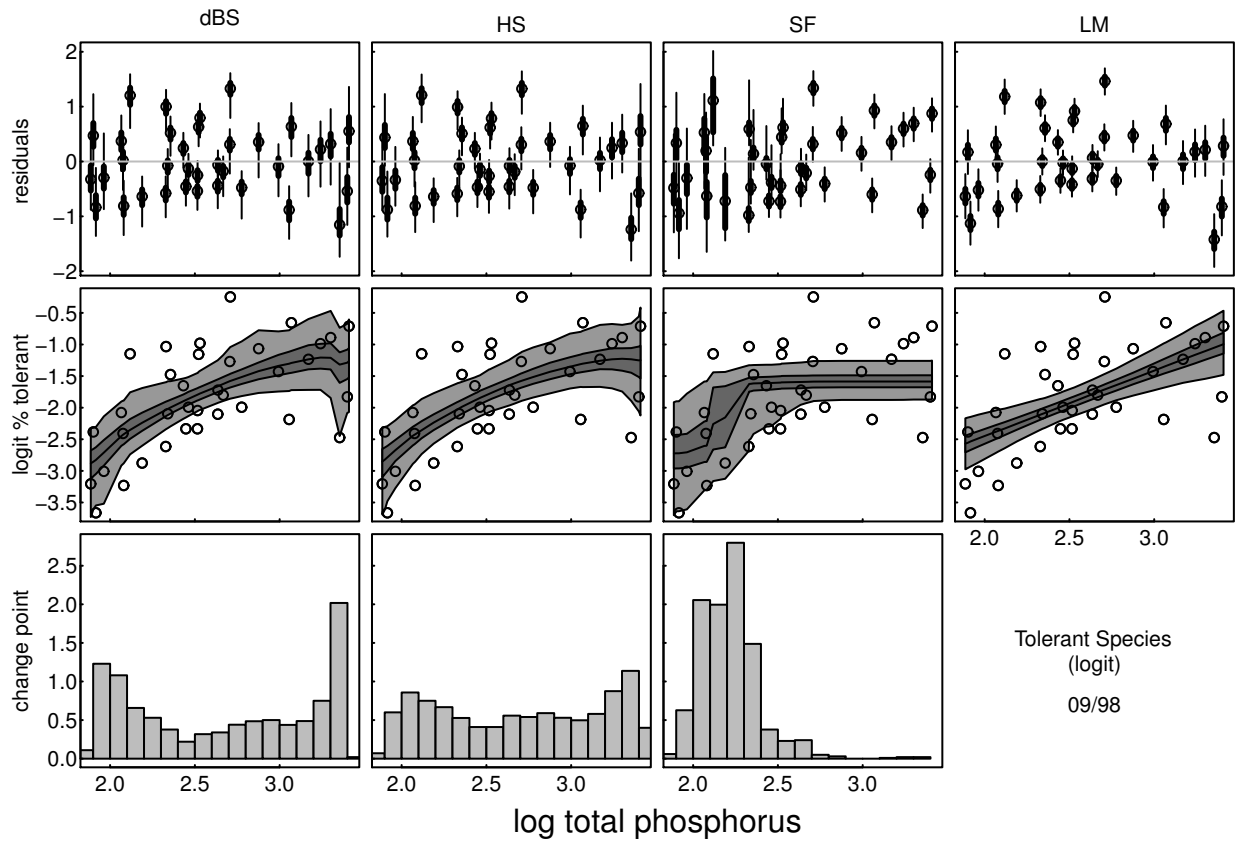


Figure 8: Logit transformed % tolerant individual data (September 1998) show that a slightly nonlinear model is the best model. Residual variances in Figures 6 to 8 are more stable residual variances compared to the same in Figures 3 to 5.

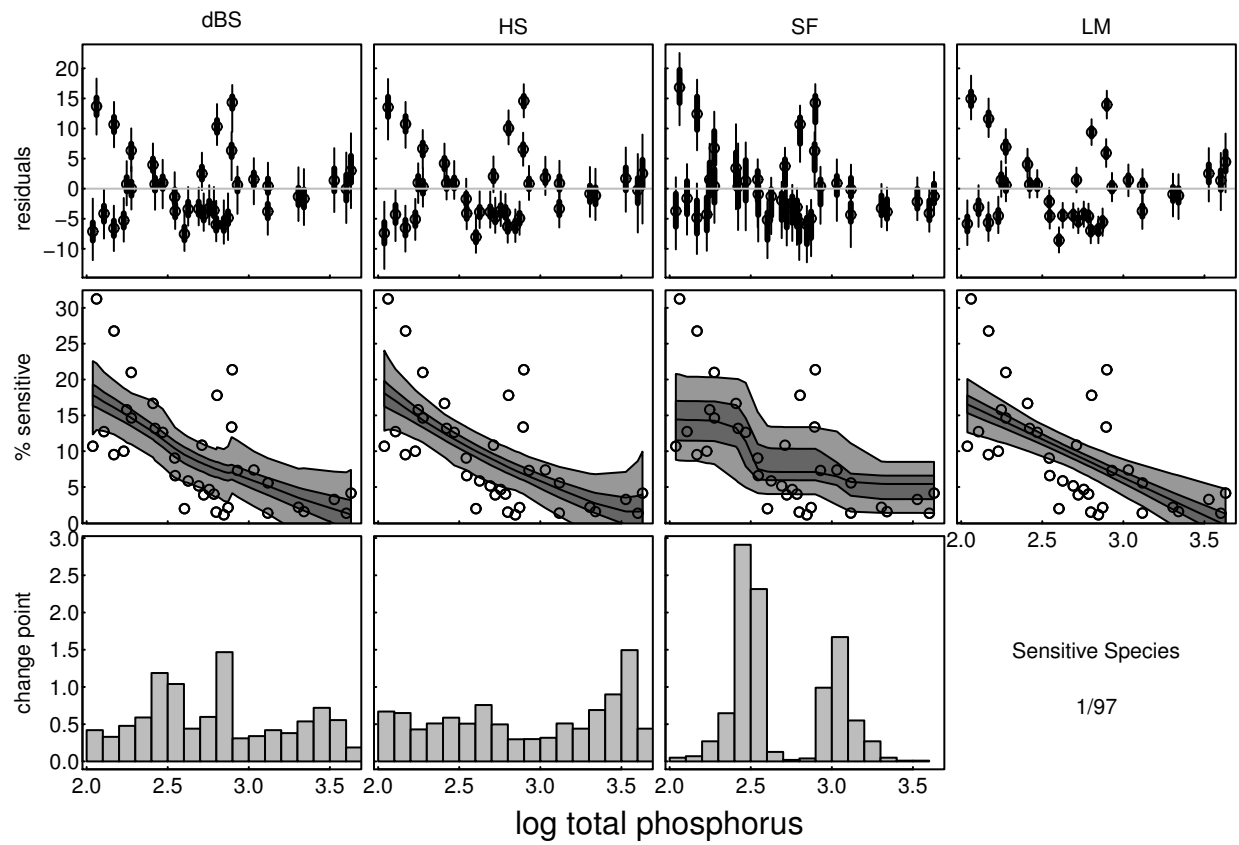


Figure 9: The response of sensitive taxa (% macroinvertebrate taxa that are sensitive to pollution defined by Florida Department of Environmental Protection, most likely Class 1 organisms defined in Terrell [1991]) to increased TP is unlikely to be a step function.

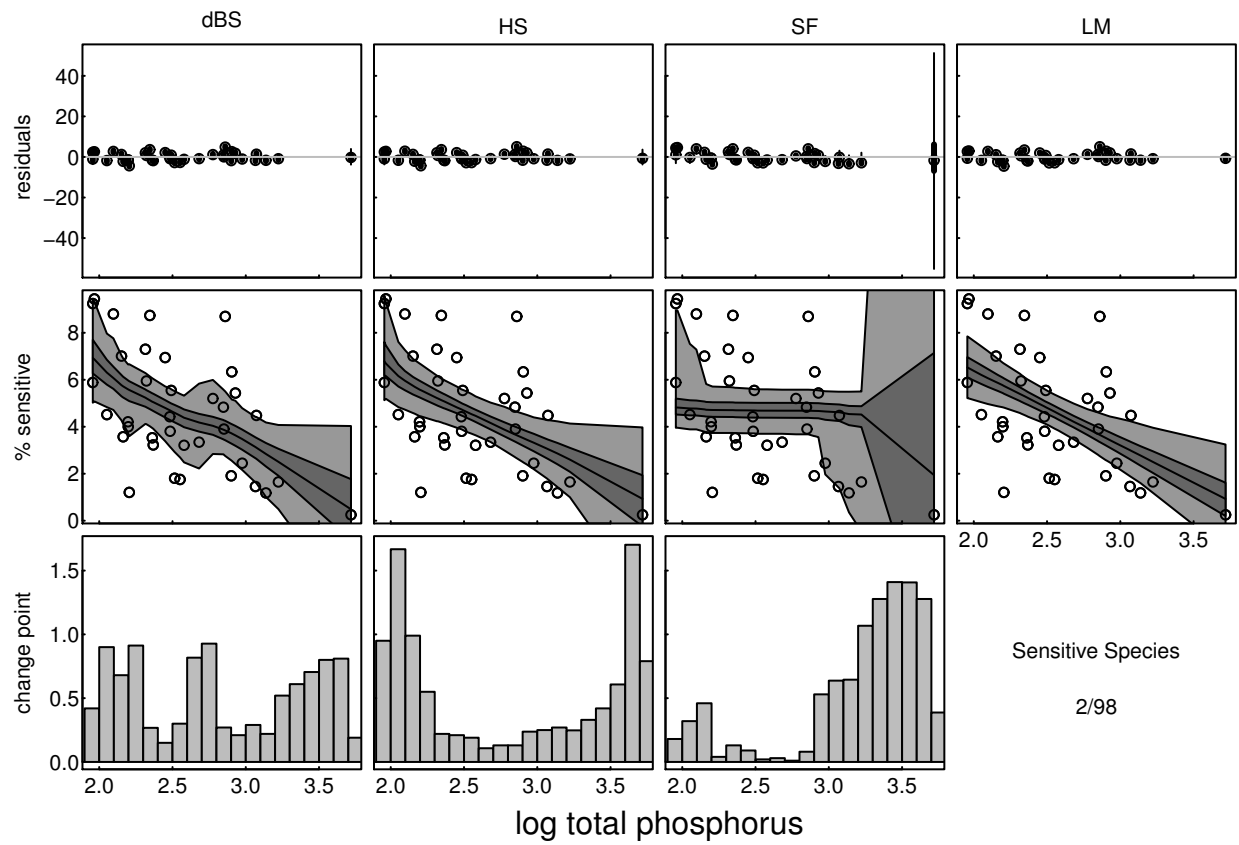


Figure 10: Sensitive species data from February 1998 are very noisy and cannot be adequately modeled by the step function model or any of the other three alternatives.

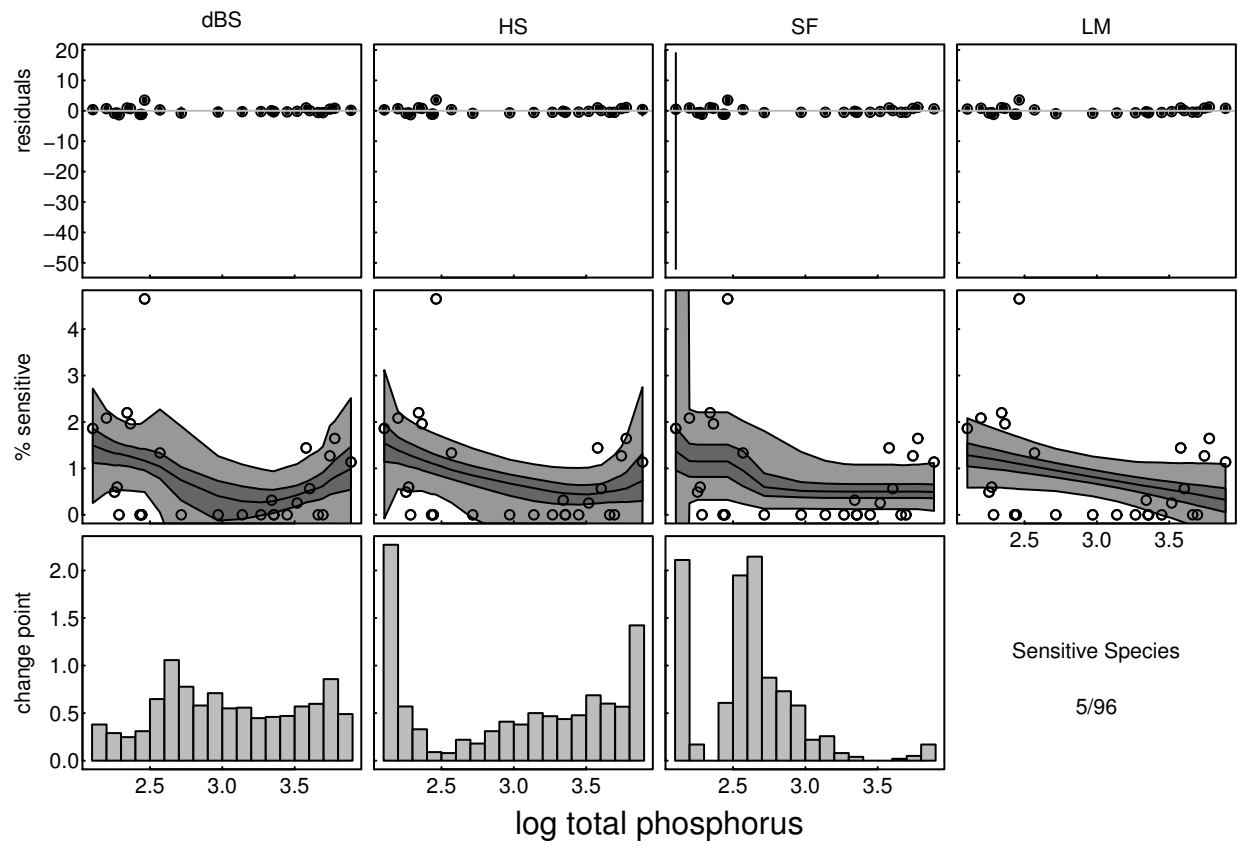


Figure 11: Sensitive species data from May 1996 are noisy and none of the four alternatives are suitable.

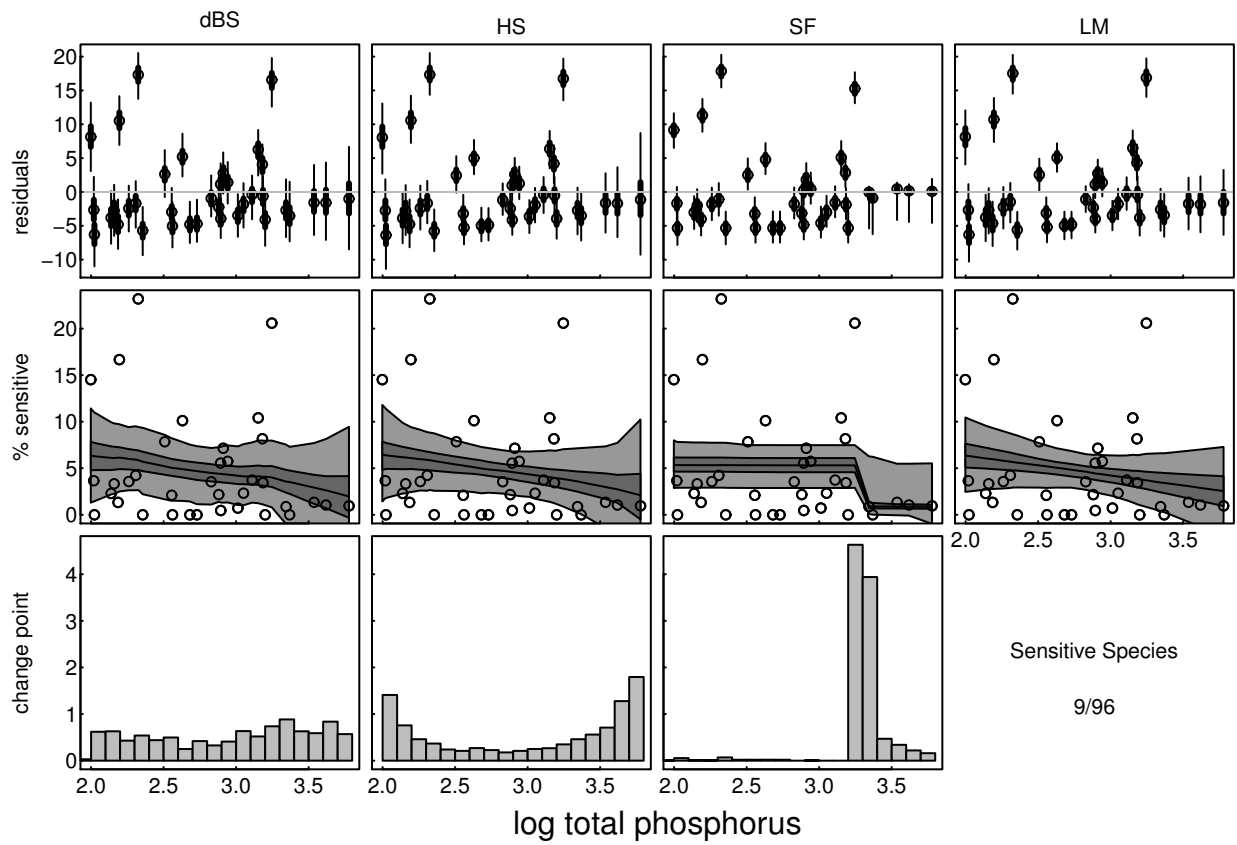


Figure 12: Sensitive species data collected from September 1996 show no relationship between % sensitive macroinvertebrate taxa and TP. As in the % tolerant individual metric, I used the logit transformation in the next four plots.

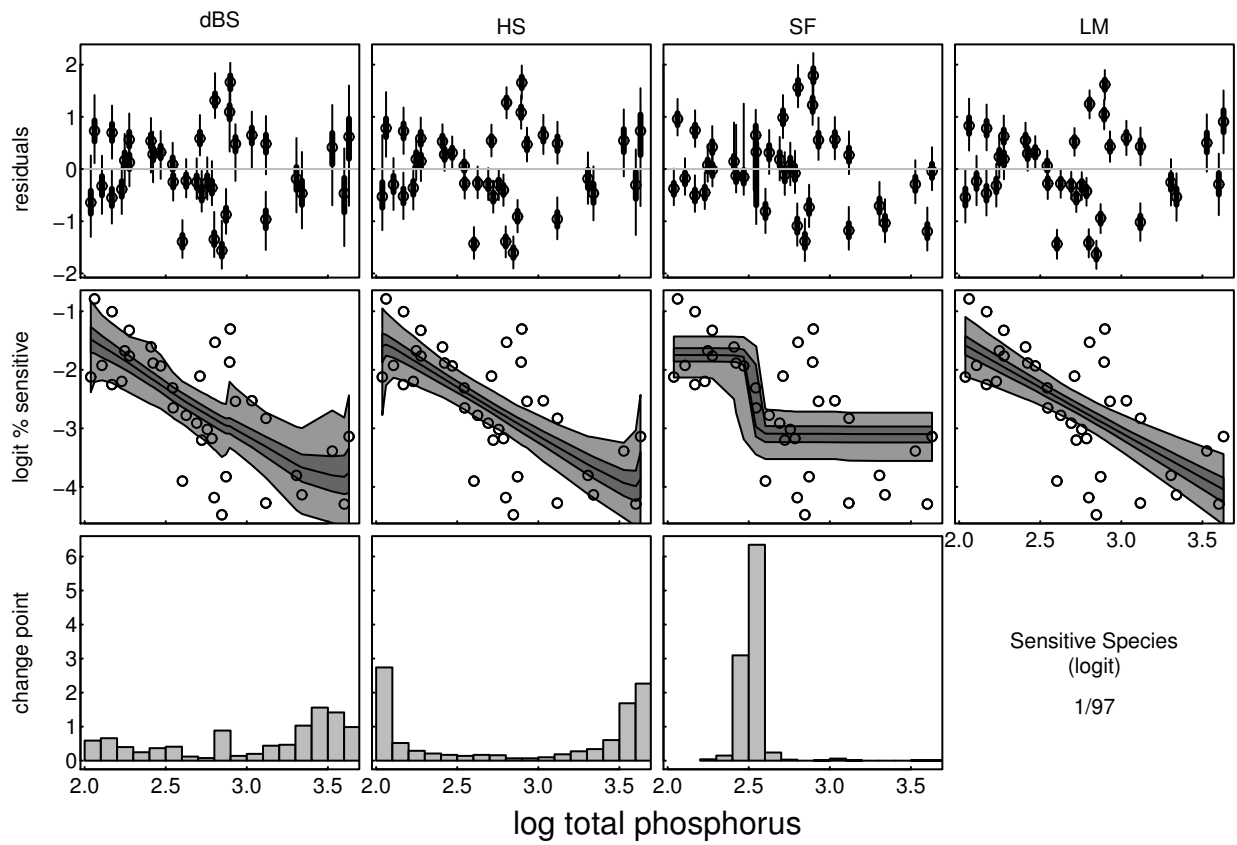


Figure 13: A linear model is probably the most appropriate model for describing the relationship between logit transformed % sensitive taxa and log total phosphorus. Data collected in January 1997. Should we use the SF model only, we may conclude the existence of a change point.

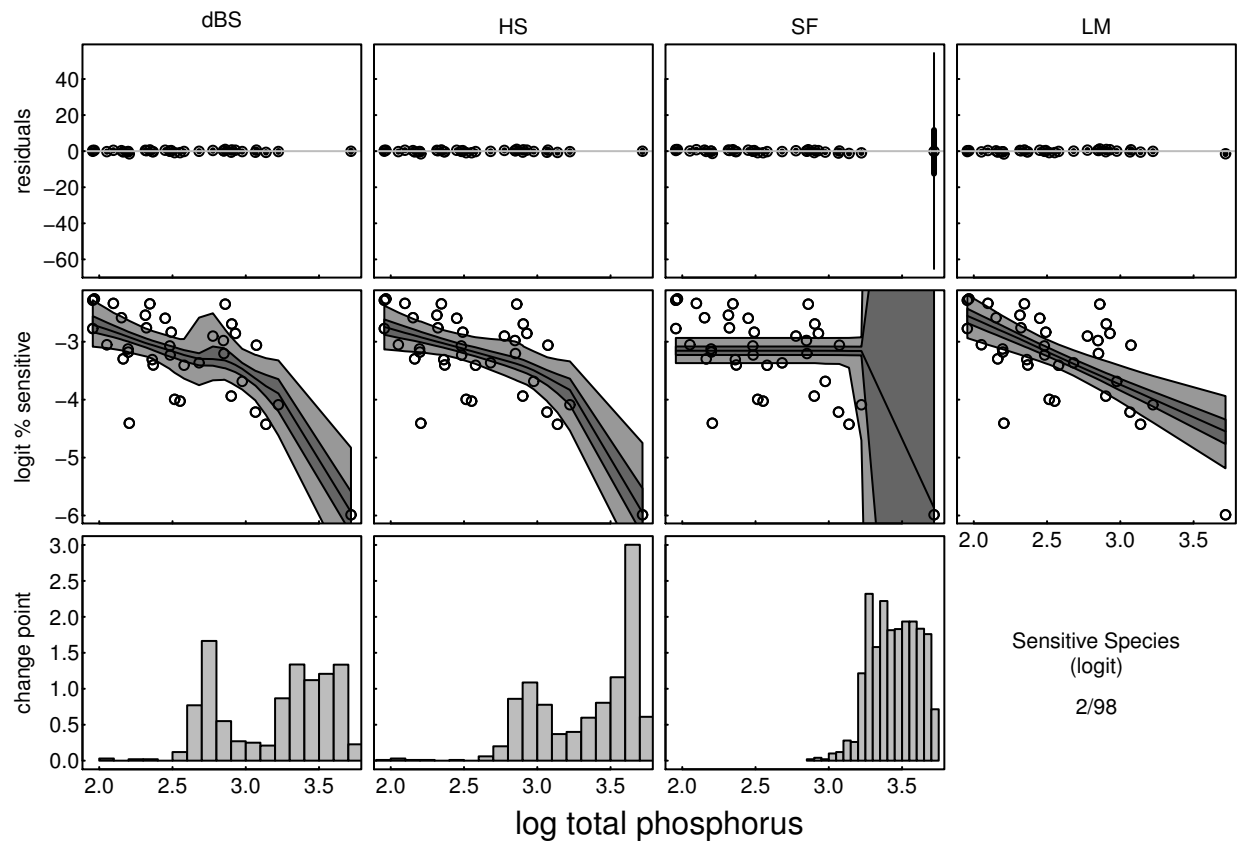


Figure 14: After logit transforming the % sensitive taxa, the data (February 1998) show little sign of a change point (of any kind).

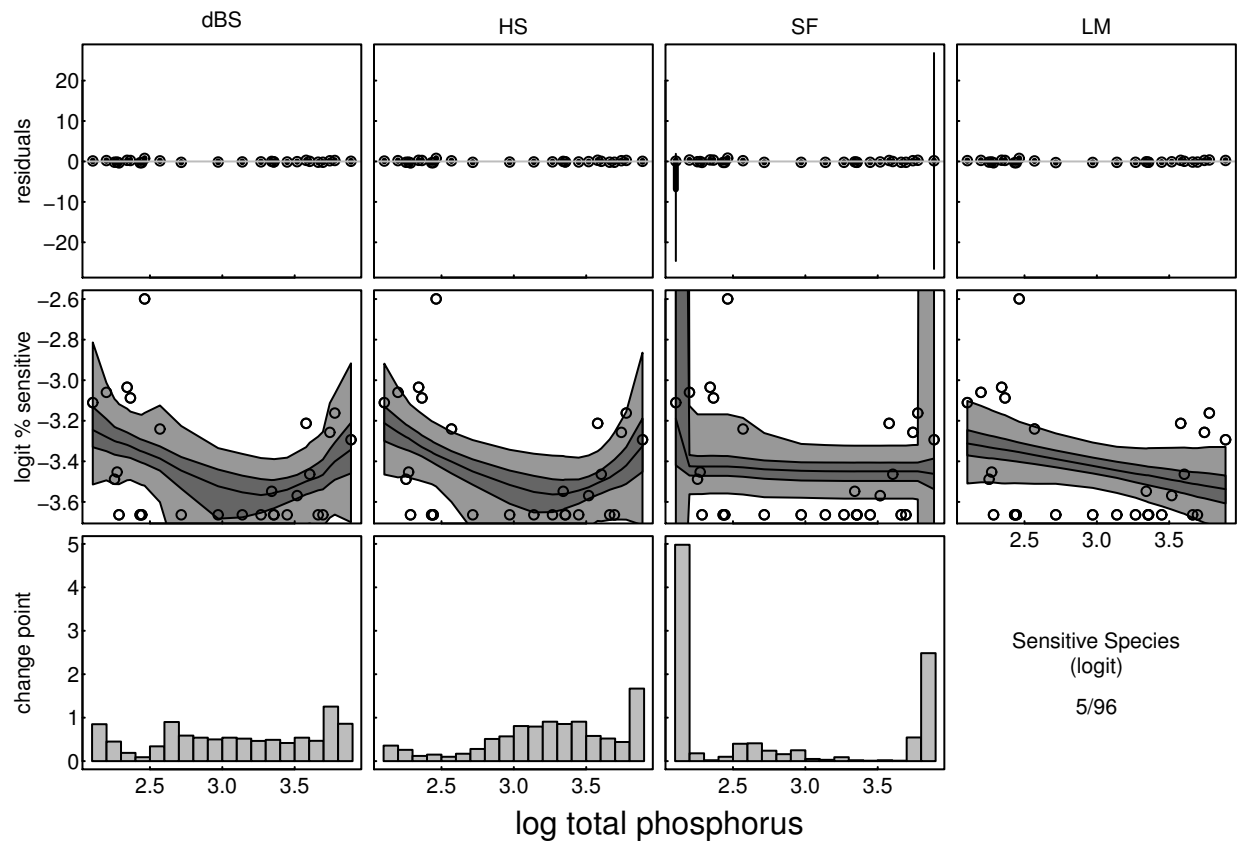


Figure 15: The logit transformed % sensitive taxa show no threshold response to increased total phosphorus.

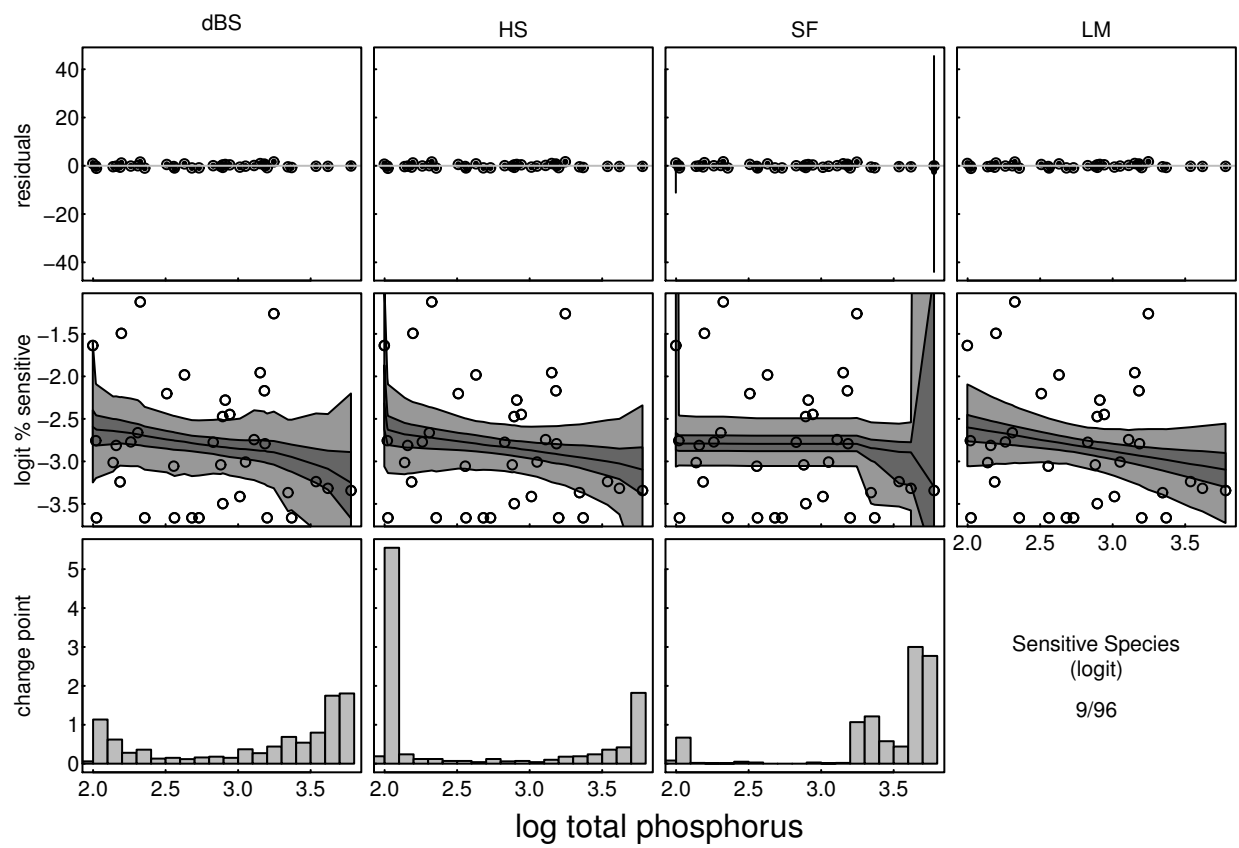


Figure 16: Sensitive taxa data collected in September 1996 show no threshold response.

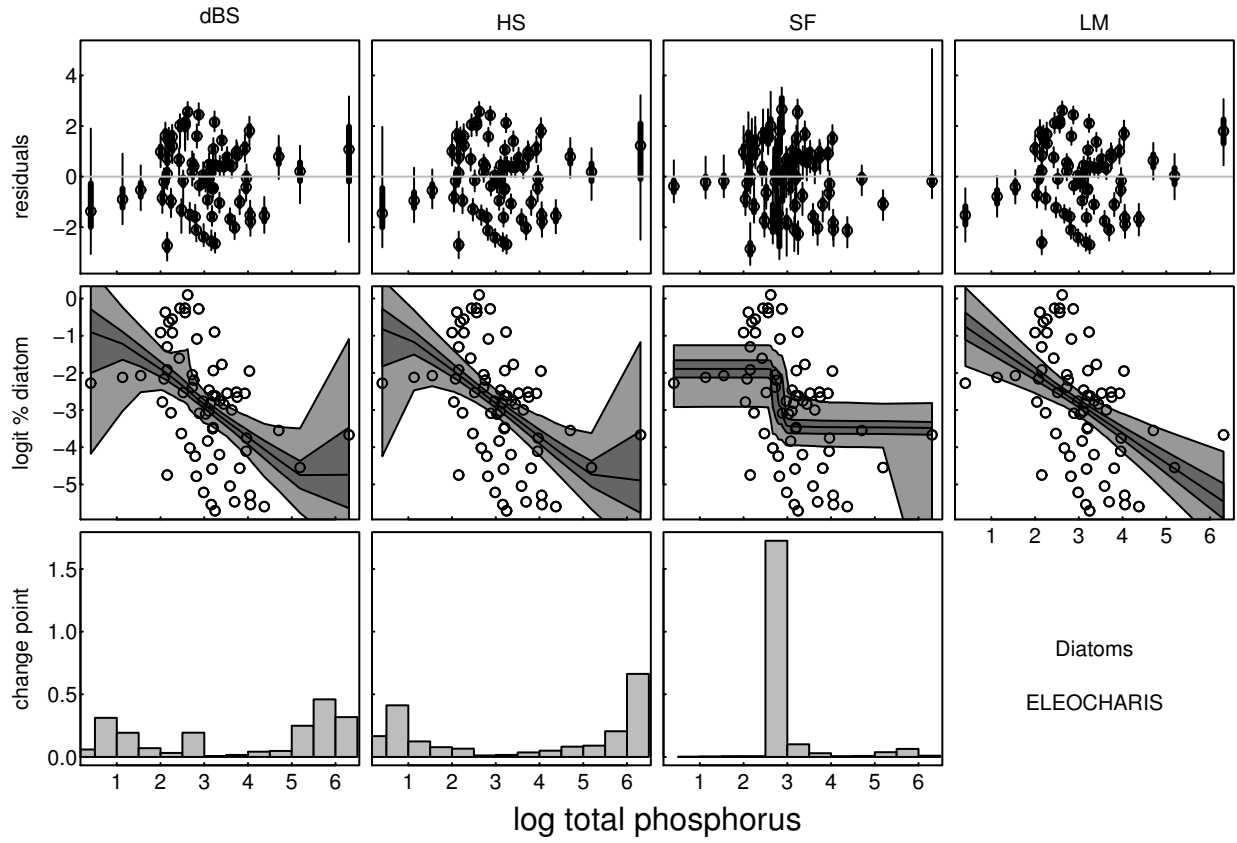


Figure 17: Logit transformed % Diatom in algae samples collected from *Eleocharis* stems shows a strong declining trend as total phosphorus increases, but the support for a change point is unavailable. Comparing the posterior change point distributions for the SF model and the broken stick model, we can confidently conclude that SF model is not appropriate.

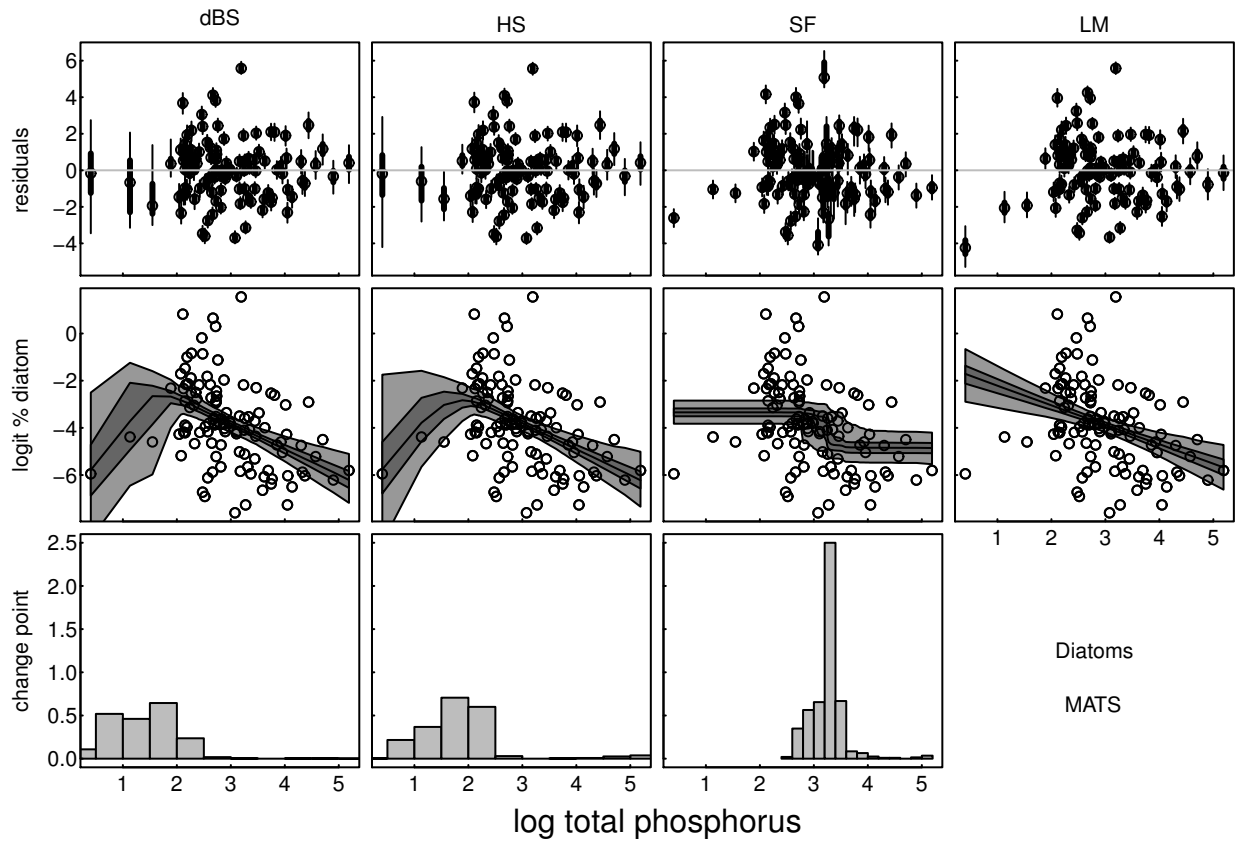


Figure 18: Same as in Figure 17, % diatom collected from floating mat does not support a change point model, although the SF posterior change point distribution is well defined and concentrated.

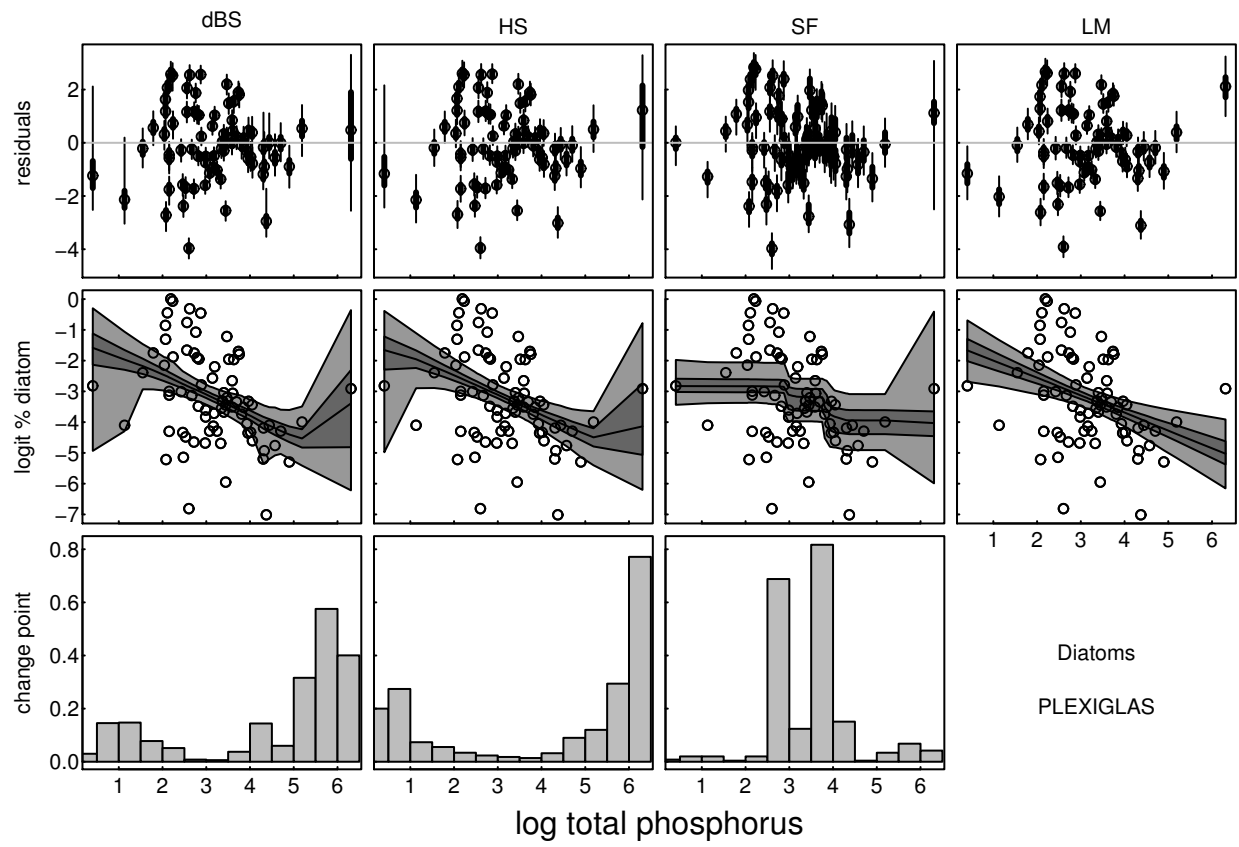


Figure 19: The evidence of a threshold response is weak in % diatom collected from plexiglass.

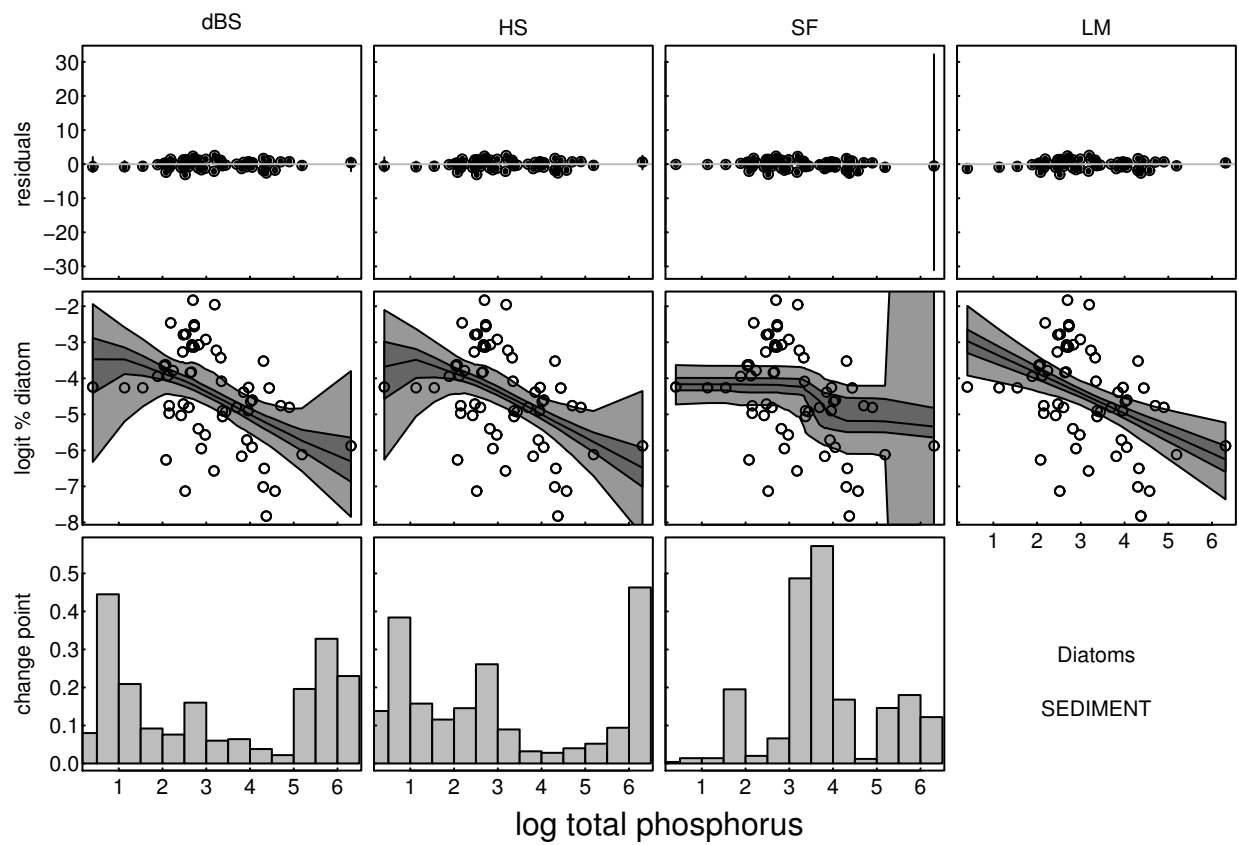


Figure 20: Data collected from sediment show no support for a change point model.

4 Residual Plots

Because the models are based on the normality assumption, assessing model residual distribution is essential. In the section, I present residual normal Q-Q plots. Because model parameters are estimated using MCMC, we have as many sets of residuals as the MCMC samples. The normal Q-Q plots presented here are generated based on the residual quantile distributions for selected quantiles (0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.975, and 0.99). Plots for the % tolerant and % sensitive taxa models are organized for easy comparison. In these Q-Q plots, the x-axis is the standard normal quantile and the y-axis is the residual quantile. Not all Q-Q plots are meaningful, especially when a model is deemed as inappropriate based on evidence illustrated in plots shown earlier in this document. For example, the three threshold models (and the linear model) are probably inappropriate for the % sensitive data. As a result, residual Q-Q plots for these models should be disregarded.

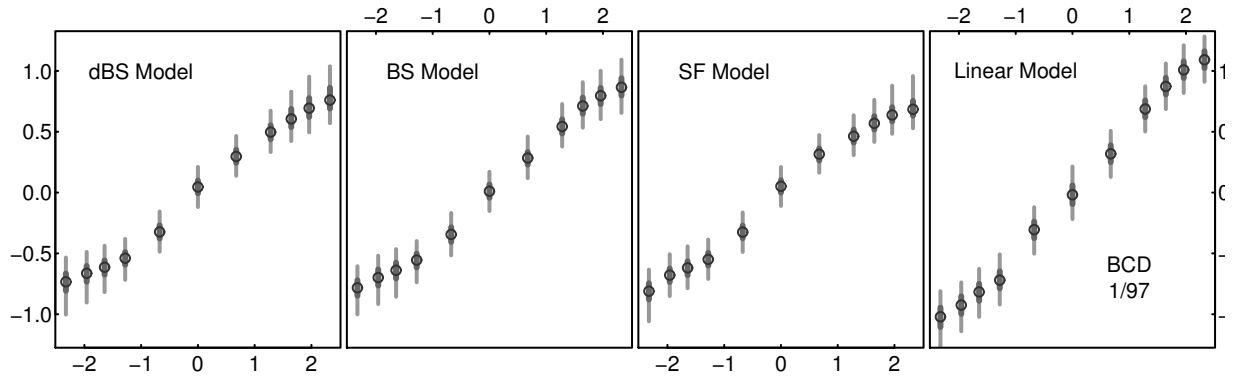


Figure 21: Residual normal Q-Q plot for the model using BCD data collected in January 1997.

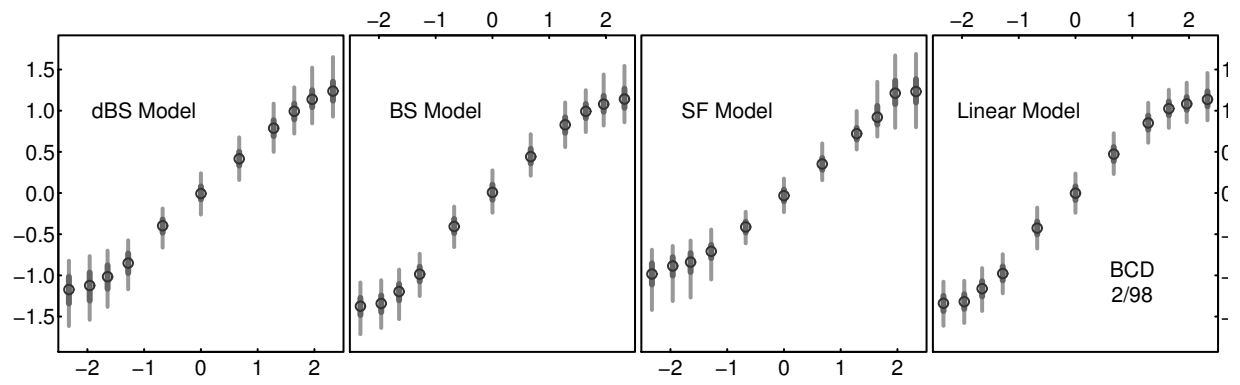


Figure 22: Residual normal Q-Q plot (BCD, February 1998).

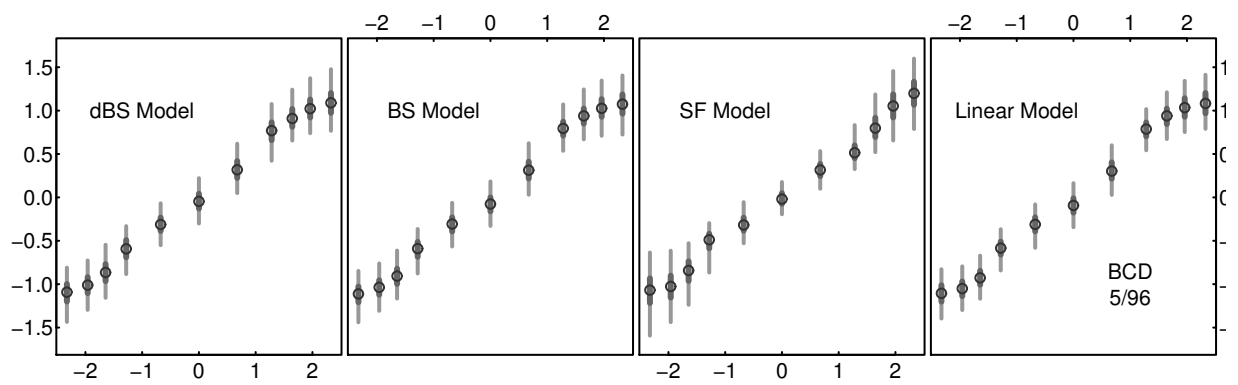


Figure 23: Residual normal Q-Q plot (BCD, May 1996).

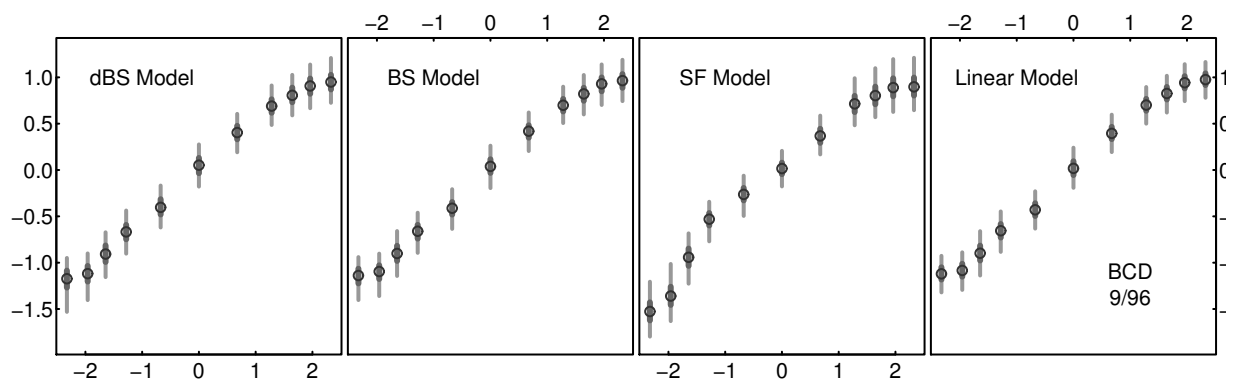


Figure 24: Residual normal Q-Q plot (BCD, September 1996).

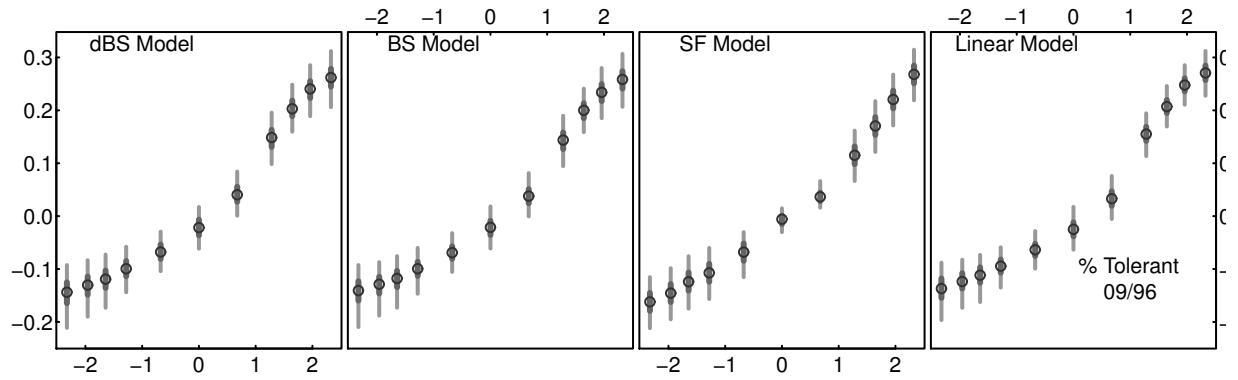


Figure 25: Residual normal Q-Q plot (% tolerant individuals, September 1996).

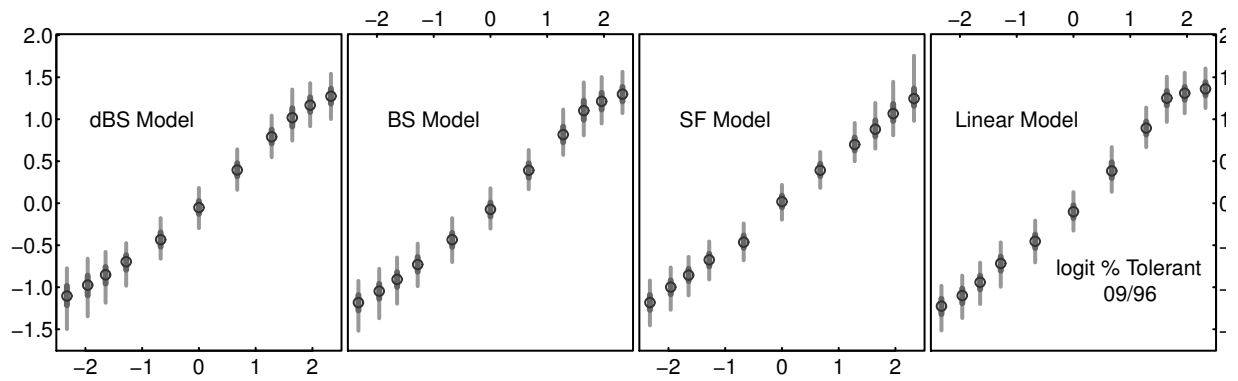


Figure 26: Residual normal Q-Q plot (logit transformed % tolerant individuals, September 1996)

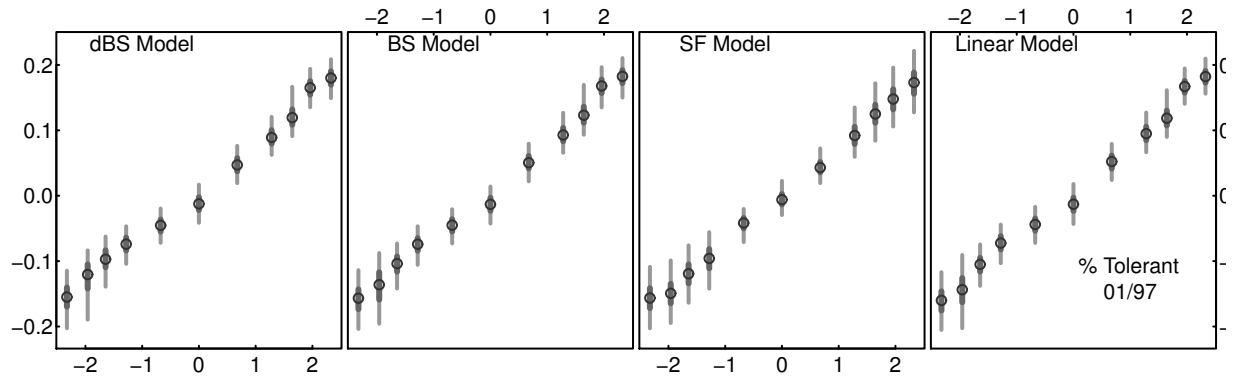


Figure 27: Residual normal Q-Q plot (% tolerant individuals, January 1997).

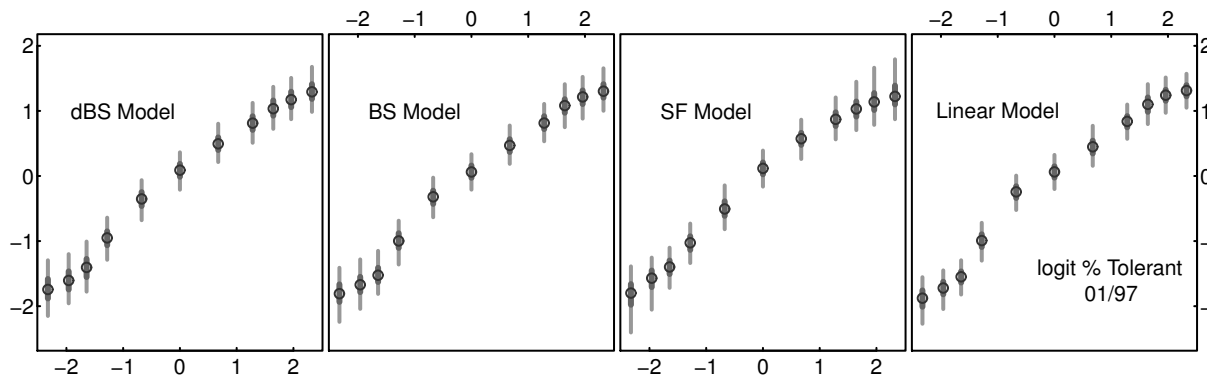


Figure 28: Residual normal Q-Q plot (logit transformed % tolerant individual, January 1997)

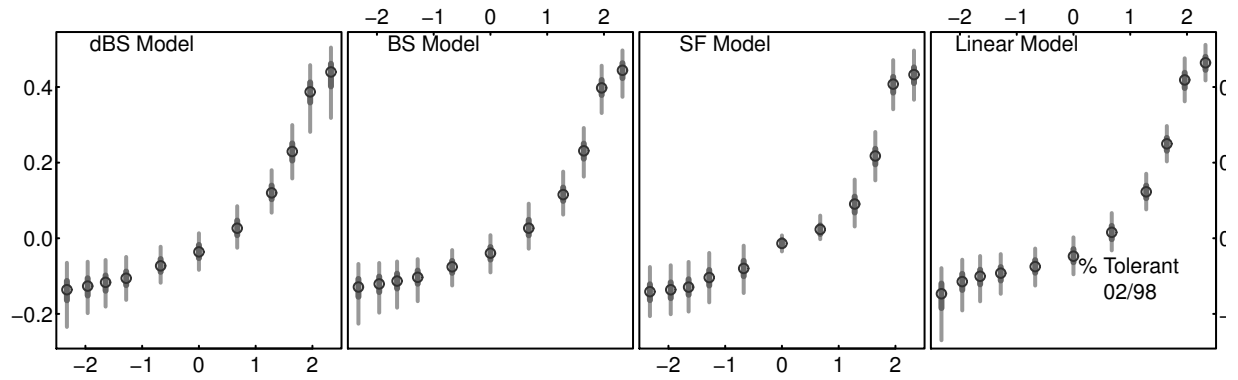


Figure 29: Residual normal Q-Q plot (% tolerant individuals, February 1998).

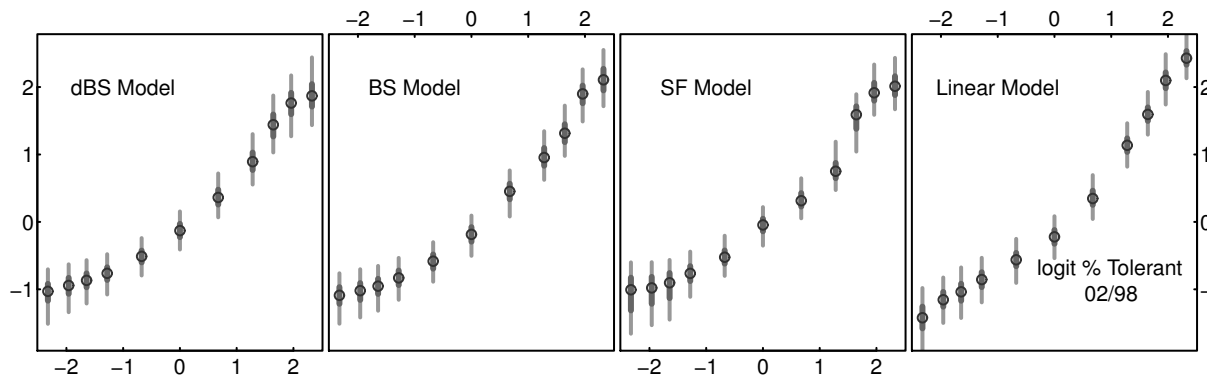


Figure 30: Residual normal Q-Q plot (logit transformed % tolerant individuals, February 1998)

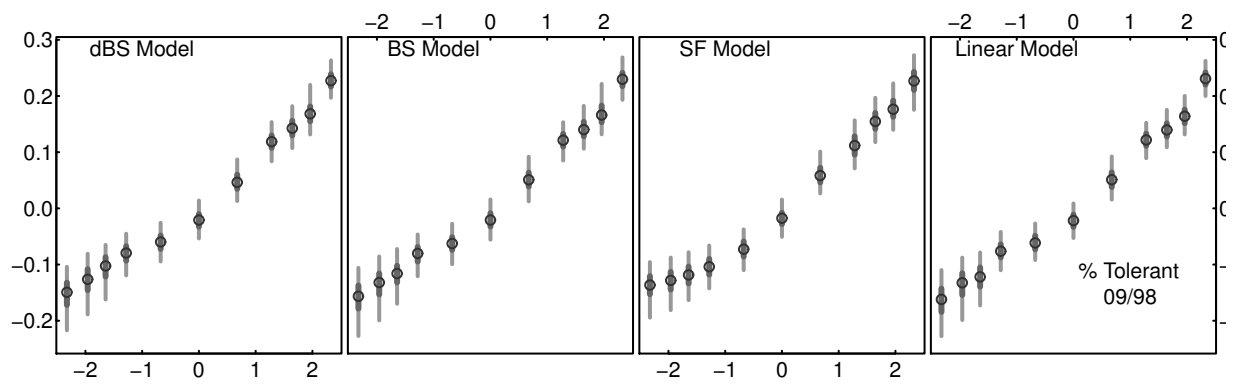


Figure 31: Residual normal Q-Q plot (% tolerant individuals, September 1998).

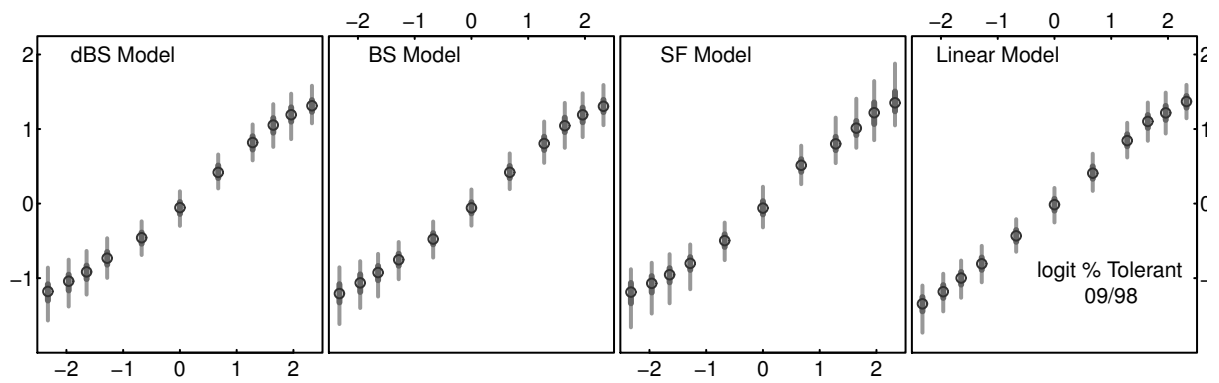


Figure 32: Residual normal Q-Q plot (logit transformed % tolerant individuals, September 1998)

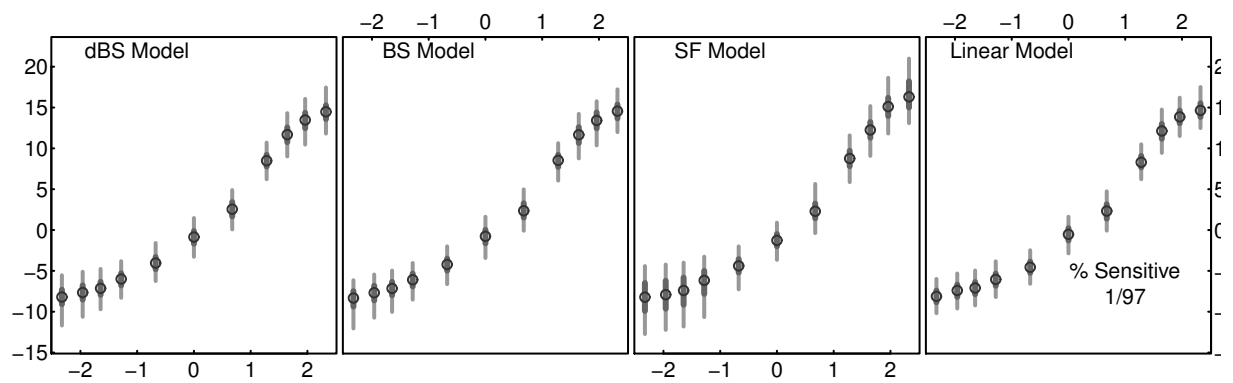


Figure 33: Residual normal Q-Q plot (% sensitive taxa, January 1997)

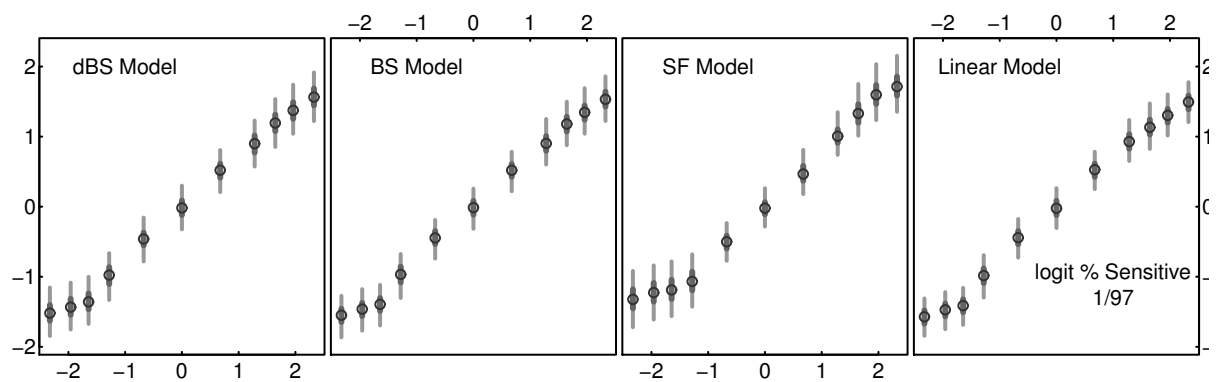


Figure 34: Residual normal Q-Q plot (logit transformed % sensitive taxa, January 1997)

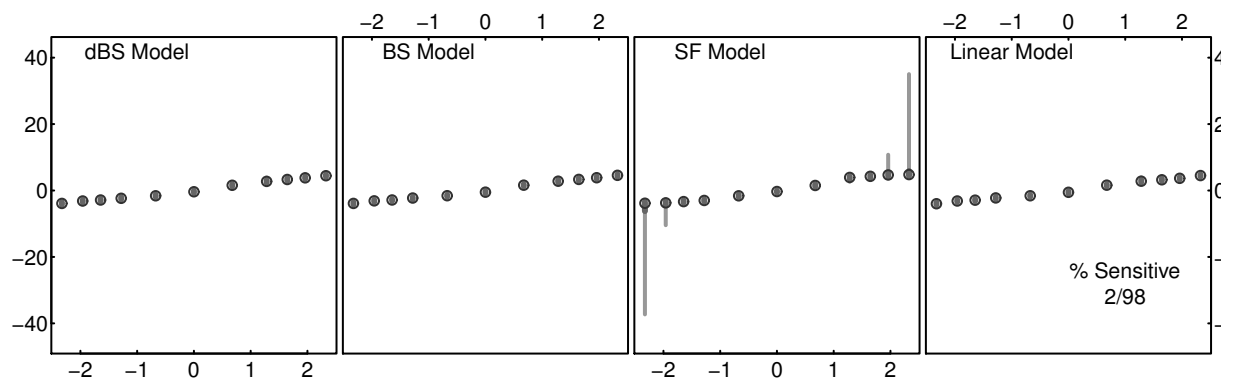


Figure 35: Residual normal Q-Q plot (% sensitive taxa, February 1998)

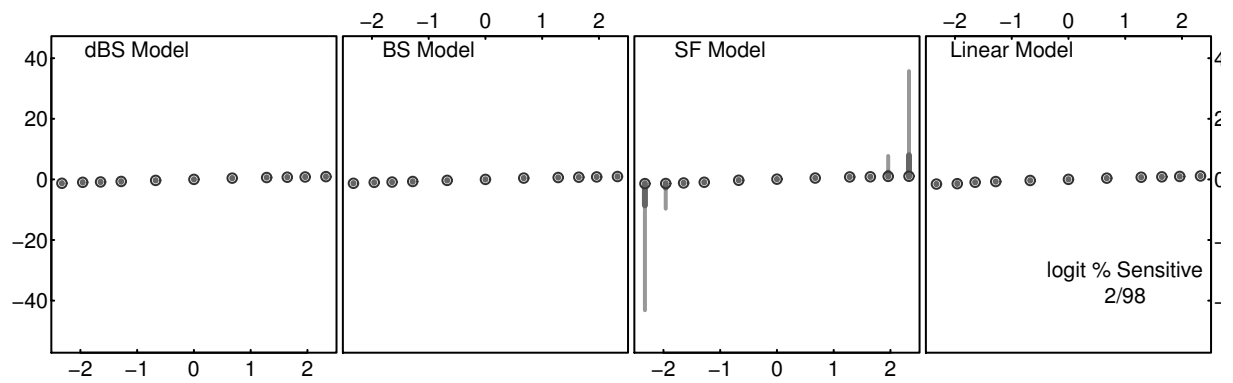


Figure 36: Residual normal quantile plot (logit transforming % sensitive taxa, February 1998)

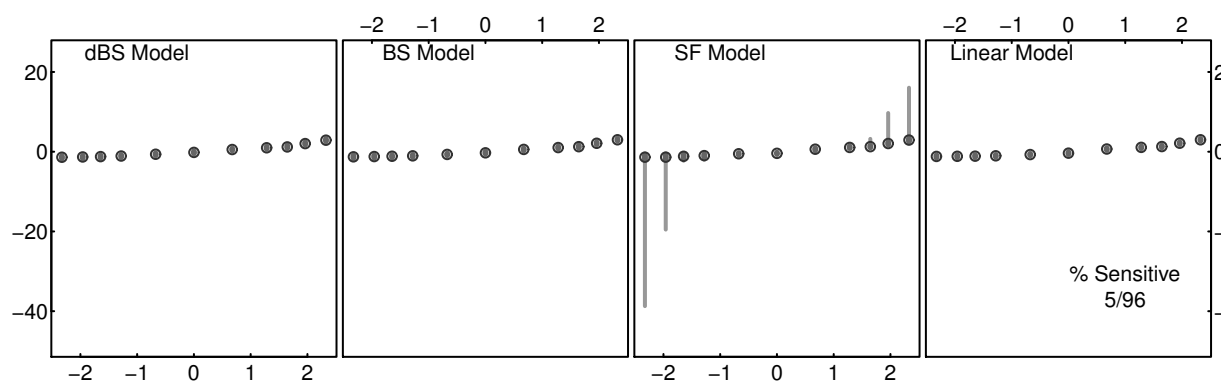


Figure 37: Residual normal Q-Q plot (% sensitive taxa, May 1996)

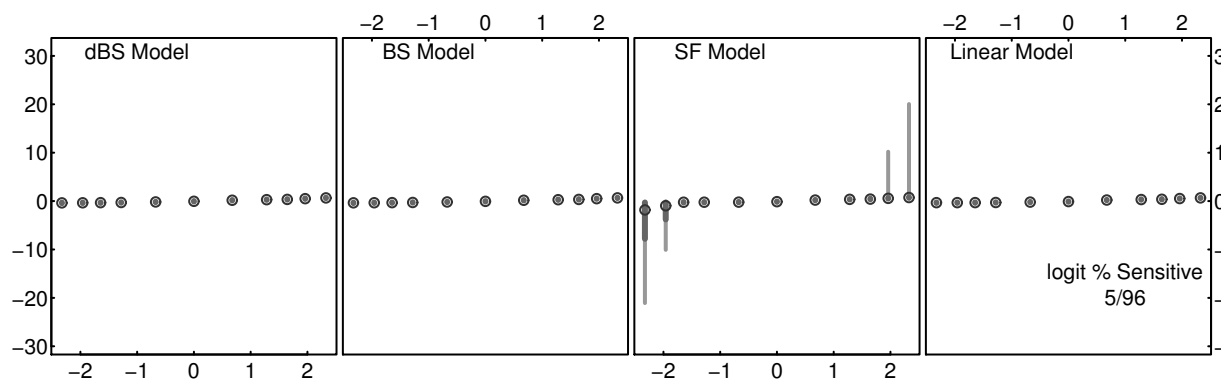


Figure 38: Residual normal Q-Q plot (logit transformed % sensitive taxa, May 1996)

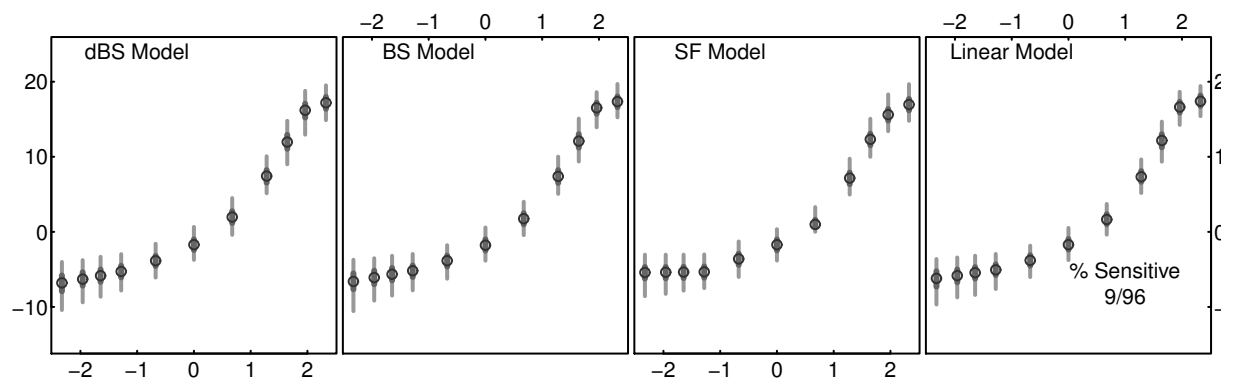


Figure 39: Residual normal Q-Q plot (% sensitive taxa, September 1996)

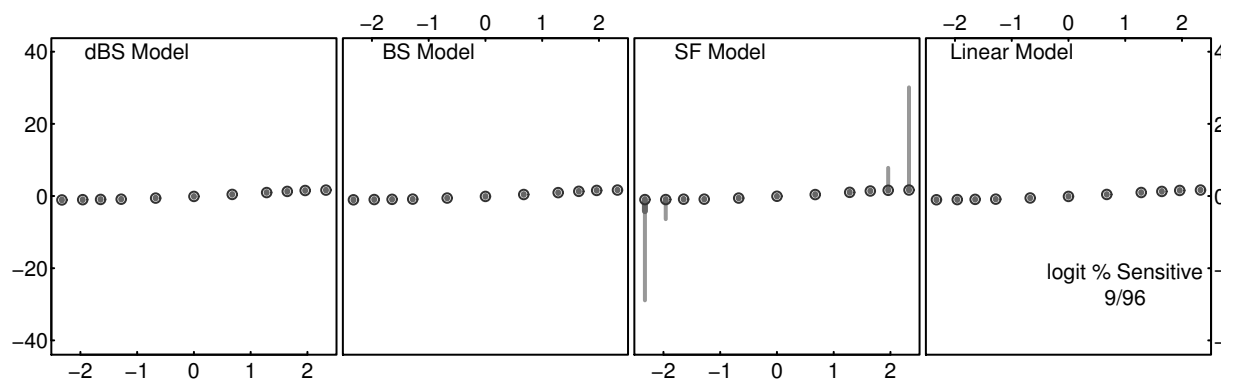


Figure 40: Residual normal Q-Q plot (logit transformed % sensitive taxa, September 1996)

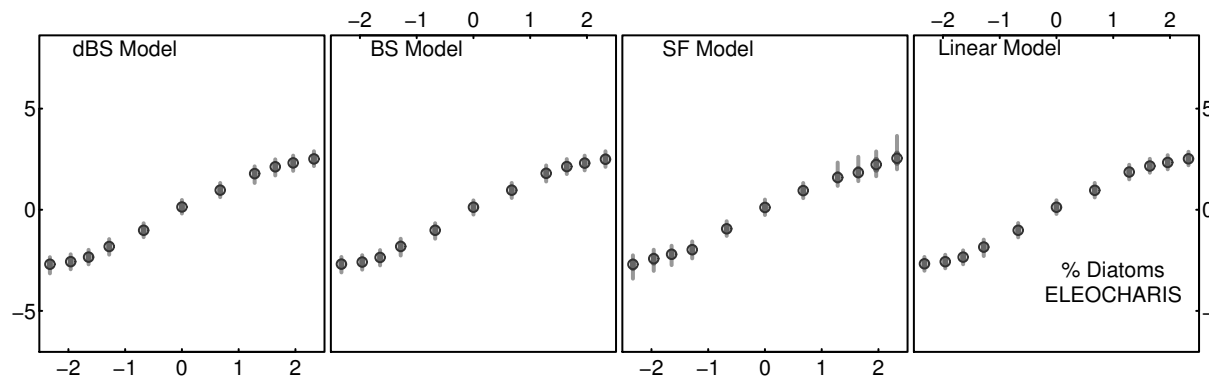


Figure 41: Residual normal Q-Q plot (logit transformed % diatom in algae samples collected from *Eleocharis* stems)

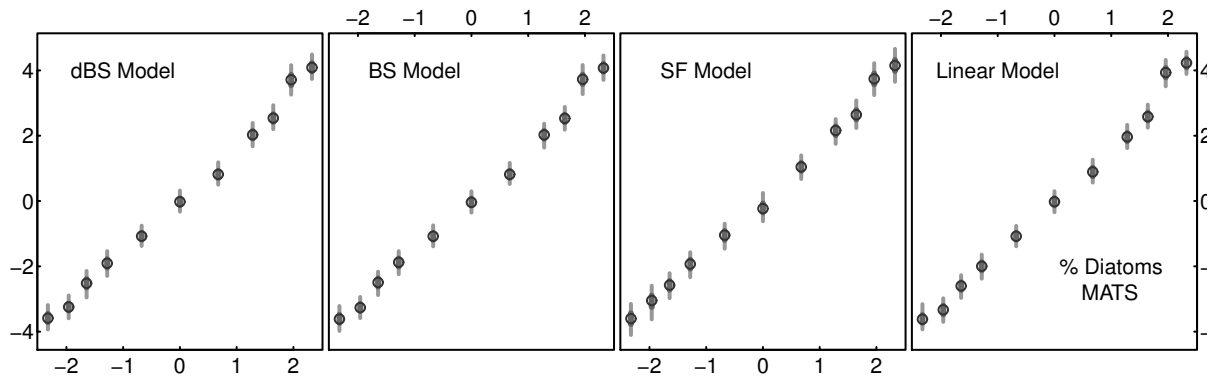


Figure 42: Residual normal Q-Q plot (logit transformed % diatom, floating mat)

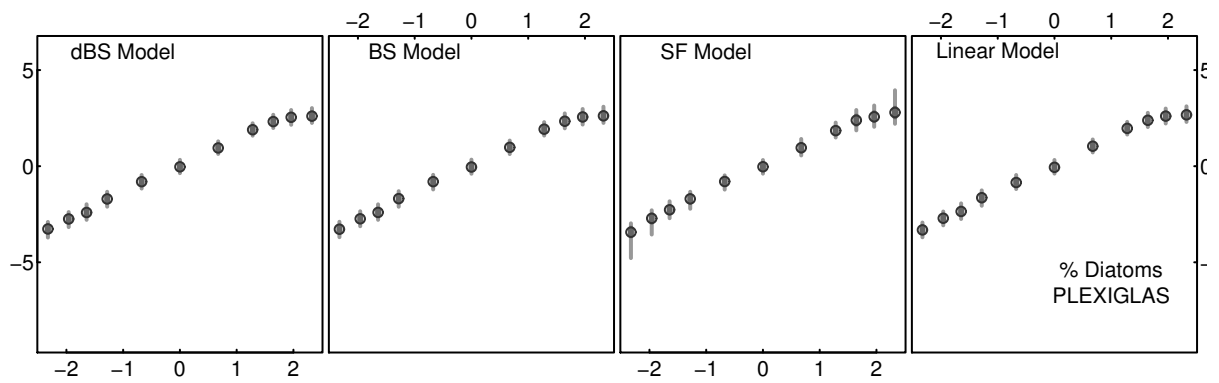


Figure 43: Residual normal Q-Q plot (logit transformed % diatom, plexiglass)

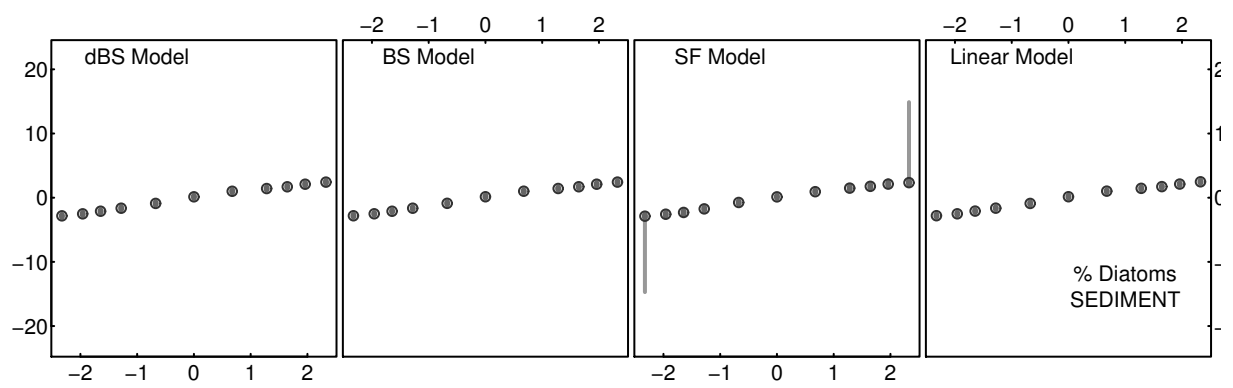


Figure 44: Residual normal Q-Q plot (logit transformed % diatom, sediment)

References

C.R. Terrell. *Water Quality Indicators Guide: Surface Waters*. Diane Publishing Company, 1991. ISBN 9781568061825.