The Lake State model and Lake Load Response internet tool

The Lake Load Response (LLR) internet tool has been developed to ease the use of models in making predictions about the effects of nutrient loading into a lake. It is based on the LakeState (LS) model, that consists of three component models: Chapra's (1975) model for retention of total phosphorus and nitrogen, the hierarchical, linear regression model for chlorophyll-*a* (Malve 2007) and the logistic regression model for phytoplankton biomass (Kauppila P., Lepistö L., Malve O. & Raateland A. Unpublished). In the LS model the mechanistic and statistic approach are combined using Bayesian inference with Markov Chain Monte Carlo (MCMC) simulation methods. This way, predictions about the water quality as well as about the model error can be made on a statistical basis, which gives more confidence into lake management planning.

With Chapra's retention model (Equation 1) it is possible to estimate the in-lake nutrient concentrations as a function of incoming loading and water outflow. The estimates are based on the assumption of equilibrium ($\frac{dC}{dt} = 0$), according to which the in-lake nutrient concentration does not change during the retention period. Especially the lake's internal loading can affect this equilibrium and weaken the reliability of the estimates.

$$V\frac{dC}{dt} = W - QC - v_s AC = 0 \qquad (1)$$
$$\Rightarrow C = \frac{W}{Q + v_s A}$$

V	lake volume (m^3)
$\frac{dC}{dt}$	change in nutrient concentration in time unit
W	external loading (mg d^{-1})
Q	outflow (m^3 d^{-1})
C	in-lake nutrient concentration (mg m^{-3})
v_s	Chapra's sedimentation rate(m d^{-1})
A	lake area (m^2)

Because of the use of Bayesian inference, Chapra's retention model is actually used to calculate the expectation of the nutrient concentration and the estimated nutrient concentration is thought to be normally distributed (Equation 2). Chapra's sedimentation rate is also thought to be normally distributed (Equation 3). The sedimentation rate is first determined using observations of the other variables, after which the model can be used to calculate concentration estimates with different loadings.

$$C \sim N(\mu, \tau^2) \qquad (2)$$

μ	expectation for in-lake nutrient concentration (mg m^{-3})
τ^2	error variance for the model

$$v_s \sim N(\mu_s, \sigma_s^2) \qquad (3)$$

μ_s	expectation for sedimentation rate (m d^{-1})
σ_s^2	variance for sedimentation rate

If there are not enough observations from the study lake, the Chapra's sedimentation rate can be estimated with regression models based on data from different kinds of lakes. The regression

models of Dillon and Rigler's (1974) trapping parameter for phosphorus (Equation 4) and nitrogen (Equation 5) are fitted to total nutrient balance data from 12 Finnish lakes. Canfield and Bachmann's regression model for Vollenweider's sedimentation rate (Equation 6) (Reckhow 1988) is fitted to total phosphorus balance data from 723 North American and European lakes. Reckhow's regression model for Vollenweider's sedimentation rate (Equation 7) (Reckhow 1988) is fitted to total nutrient balance data from 70 lakes and reservoirs of the southeastern parts of North-America. Chapra's sedimentation rate can then be calculated (Equations 8 and 9) using the trapping parameter and sedimentation rate values obtained from the models.

$$R_P = 0.60 - 0.0006W \qquad (R^2 = 0.47) \tag{4}$$

$$R_N = 0.35 - 0.00016T_W \qquad (R^2 = 0.48) \tag{5}$$

$$\sigma = 0.129(C_{in})^{0.549}T_W^{-0.549} \tag{6}$$

$$\sigma = \beta C_{in}^{\theta 1} T_W^{\theta 2} z^{\theta 3} \tag{7}$$

$$v_s = \frac{R_{P/N}}{1 - R_{P/N}} q_s \tag{8}$$

$$R_{P/N} = (1 + \frac{1}{T_W \sigma})^{-1} \tag{9}$$

R_P	phosphorus retention coefficient
W	external loading
R_N	nitrogen retention coefficient
T_W	hydraulic detention time
σ	sedimentation coefficient (a^{-1})
C_{in}	nutrient concentration in inflowing water (mg m^{-3})
β	for phosphorus = 3.0 (SD ± 0.25), for nitrogen = 0.67 (SD ± 0.10)
$\theta 1$	for phosphorus = 0.53 (SD ± 0.13), for nitrogen = 0
$\theta 2$	for phosphorus = -0.75 (SD ± 0.06), for nitrogen = -0.75 (SD ± 0.11)
$\theta 3$	for phosphorus = 0.58 (SD ± 0.19), for nitrogen = 0
z	lake mean depth (m)
v_s	Chapra's sedimentation rate (m d^{-1})
q_s	hydraulic surface loading (m^3 a^{-1})

LLR gives water quality predictions using all the alternative ways of calculating the sedimentation rate. The estimate of the lake specific model is based only on observations from the lake. The "Finnish lakes" -model calculates the estimates using the sedimentation rates obtained from the data of Finnish lakes. The "North American and European lakes" -model uses the sedimentation rates obtained from the data of North American and European lakes, as well as lakes and reservoirs of the southeastern parts of North-America. Predictions of the lake specific model are usually most reliable, because they are based on real observations from the study lake. However, if there are not enough observations from the lake, or if the observation values cover a very narrow range, the other two models can help to increase the reliability of the predictions. There may still be problems with how well regressions drawn from lakes from different areas or from a different kind of lake fit a certain lake.

The in-lake phosphorus and nitrogen concentrations can be used to predict the in-lake chlorophyll-a concentration with the hierarchical, linear regression model for chlorophyll-a (Malve & Qian 2006, Lamon et al 2008). The model may be summarized as follows (Malve 2007):

$$\log(y_{ijk}) \sim N(X\beta_{ij}, \tau^2) \tag{10}$$

$$X\beta_{ij} = \beta_{0,ij} + \beta_{1,ij}\log(TP_{ijk}) + \beta_{2,ij}\log(TN_{ijk}) + \beta_{3,ij}\log(TP_{ijk})\log(TN_{ijk})$$

$$\beta_{ij} \sim N(\beta_i, \sigma^2_i)$$

$$\beta_i \sim N(\beta, \sigma^2)$$

$\log(y_{ijk})$	*k*th observed log(Chla) value from lake *j* of type *i*
X	matrix containing the observed total phosphorus (TP) and total nitrogen (TN) values from lake *j* of type *i*
β_{ij}	lake specific model parameter vector $[\beta_{0,ij}, \beta_{1,ij}, \beta_{2,ij}, \beta_{3,ij}]$ which consists of the intercept ($\beta_{0,ij}$) and slopes for log(TP) ($\beta_{1,ij}$), log(TN) ($\beta_{2,ij}$) and for the combined effect of log(TP) and log(TN) ($\beta_{3,ij}$)
τ^2	model error variance
β_i	vector $[\beta_{0,i}, \beta_{1,i}, \beta_{2,i}, \beta_{3,i}]$ of the model parameter means for lake type *i*
σ^2_i	vector $[\sigma^2_{0,i}, \sigma^2_{1,i}, \sigma^2_{2,i}, \sigma^2_{3,i}]$ of variances in model parameters between lakes of type *i*
β	means for lake types $[\beta_0, \beta_1, \beta_2, \beta_3]$
σ^2	variance for lake types $[\sigma^2_0, \sigma^2_1, \sigma^2_2, \sigma^2_3]$

From the relation between nutrient and chlorophyll-a concentration it is then possible to determine the relation between loading and chlorophyll-a concentration. This leads to predictions about the target load with which a good water quality according to chlorophyll-a concentration is achieved. The hierarchy of the model means that it uses both the data from the study lake and from the lakes of same type to make the predictions. The lake type specific data, that includes observations from 2000 Finnish lakes, is already in the LLR database. The main basis for the usage of the hierarchical model is that lakes within the same type are assumed to have a similar chlorophyll-a response to changing nutrient concentrations. It is also assumed that data from one lake type covers a wider range of observation values than that from a single lake.

In practice the chlorophyll-a predictions are based almost solely on the data from the study lake when there is plenty of it, or if there is a very large scatter, for any reason, in the chlorophyll-a response to nutrient concentrations within the lakes of same type. In the opposite situation the predictions are based on the lake type specific data, but usually the emphasis on different data sources is somewhere in between. The use of lake type specific data increases the reliability of the predictions, especially when the target loads are extrapolated outside the range of observational values from the study lake.

The logistic regression model for phytoplankton biomass (Equation 11) gives predictions about the probability of phytoplankton biomass to exceed the boundary of good water quality with different phosphorus and nitrogen loads. The model is fitted to two data sets that consist of observations from Finnish and Norwegian lakes. The first data set represents humic lakes (362 lakes, color > 40 mg Pt/l) and the second clear water lakes (852 lakes, color < 40 mg Pt/l). The division to two data sets is made because in clear water lakes the exceeding probability has shown to be most related to phosphorus concentration, but in humic lakes both to phosphorus and nitrogen concentration.

$$\text{Logit}(Pr) = \beta_n + \beta_1 \log(C_P) + \beta_2 \log(C_N) \qquad (11)$$

$$\beta_i \sim N(\mu_R, \sigma_R^2)$$

Pr	probability of bloom
β_n	regression intercept
β_i	regression coefficient
C_i	observed total phosphorus/ nitrogen concentration
μ_R	expectation for regression coefficients
σ_R^2	variance for regression coefficients

References

Chapra S. 1975. Comment on "An empirical method of estimating the retention of phosphorus in lakes" by W. B. Kirchner and P. J. Dillon. *Water Resour. Res.* 11: 1033-1034.

Dillon P. & Rigler F. 1974. The test of simple nutrient budget model predicting the phosphorus concentration in lake water. *Res. Bd. Can.* 31: 1771-1778.

Lamon E.C., Malve O. & Pietilainen O.-P. 2008. Lake classification to enhance prediction of eutrophication endpoints in Finnish lakes. *Environmental Modelling & Software*. 23: 938-947. doi:10.1016/j.envsoft.2007.10.008

Malve O. 2007. *Water quality prediction for river basin management*. PhD. thesis, Helsinki University on Technology. 126 s.

Malve O. & Qian S. 2006. Estimating nutrients and chlorophyll a relationships in Finnish Lakes. *Environmental Science & Technology*. 40: 7848-7853. DOI: 10.1021/es061359b

Reckhow K.H. 1988. Empirical models for trophic state in southeastern U.S. lakes and reservoirs. *Water Resour. Bull.* 24: 723-734.