SEVENTH FRAMEWORK PROGRAMME

THEME 6: Environment (including Climate Change)

REFRESH

Adaptive strategies to Mitigate the Impacts of Climate Change on European Freshwater Ecosystems

Collaborative Project (large-scale integrating project)
Grant Agreement 244121
Duration: February 1st, 2010 – January 31st, 2014

Deliverable 5.7: Report on suitable surrogates for use in catchment-scale modelling

Lead contractor: **University of Duisburg-Essen (UDE)**
Other contractors involved: **University of Reading (UREAD)**

Due date of deliverable: **Month 18**
Actual submission date: **Month 19**

Work package:	5
Contributors:	Daniel Hering, Richard Skeffington, Andrew Wade
Estimated person months:	4

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013) Dissemination Level (add X to PU, PP. RE or CO)		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract

The aim of this deliverable is to explore the options for the application of Bayesian Models in the REFRESH study catchments. Most importantly, we search for biotic target variables which are suitable for characterising each catchment, and for environmental variables which can potentially influence the biota and which could be the subject of physical modelling. These combinations of environmental and biotic variables will later be linked by the definition of "knowledge rules".

After a short introduction to Bayesian Models, we summarize here the response of partners responsible for the REFRESH model catchments on how best to link environmental and biotic variables. Based on this, we suggest target variables (i.e. biotic metrics) for the majority of catchments.

For some catchments the model concepts are already advanced. For Lake Pyhäjärvi (Finland), relationships between nutrients and phytoplankton composition are modelled in a mechanistic way. For Lake Vansjø-Hobøl (Norway), Bayesian Networks have already been constructed. For the wetlands of the Louros catchment (Greece) a detailed concept has been developed.

For the other catchments we define "candidates" for biological variables to be modelled to be used as target variables in Bayesian Networks. This concerns River Dee (Scotland), Rivers Thames and Kennet (England), Vlatava river catchment (Czech Republic) and La Tordera (Spain). Preference was given to metrics compliant to the Water Framework Directive or to "intercalibration common metrics" meeting the taxonomic resolution usually applied in the country for the organism group in question.

Aims and scope

One of the fundamental modelling tasks in REFRESH is the integration of ecosystem models for rivers, lakes and wetlands with models of key ecological vulnerability indicators to determine interactions between climate and land-use management change and freshwater ecology (Task 5.1).

Only in few cases can this be achieved with mechanistic models, as relationships between environmental variables (such as temperature, nutrients or flow) and biotic variables (such as species composition or biomass) are complex and many linkages are not well known. Alternatively, Bayesian Models can be used, which are based on known (or assumed) statistical relationships between explanatory environmental variables and a biotic target variable.

The aim of this deliverable is to explore the options for the application of Bayesian Models in the REFRESH study catchments. Most importantly, we search for biotic target variables which are suitable for characterising each catchment, and for environmental variables which can potentially influence the biota and which could be the subject of physical modelling. These combinations of environmental and biotic variables will later be linked by the definition of "knowledge rules".

After a short introduction to Bayesian Models we summarize here the response of partners responsible for the REFRESH model catchments on how best to link environmental and biotic variables. Based on this, we suggest target variables (i.e. biotic metrics) for the majority of catchments.

The term "surrogate" given in the title of this deliverable might not be ideal for describing the Bayesian approach. We understand this term here as a synonym for "target variable" in Bayesian Models.

Some considerations on the use of Bayesian Models to link environmental and biotic data

For detailed information see Charniak (1991) and Stewart-Koster et al. (2010).

In general, Bayesian Networks belong to the category of probabilistic graphical models. A Bayesian network is a graphical structure (i.e. a flowchart) visualizing the interdependency of predictors (e.g. phosphate concentration and flow diversity) and response variables (e.g. the fish assemblage). A Bayesian network is composed of nodes and edges. Nodes are variables (depending on the topic investigated); edges are probabilistic dependencies between the variables. The variable represented by a node may be continuous (e.g. a score or temperature) or divided into classes. For example: the macroinvertebrate community might be defined as ecological quality classes good, moderate, poor or absent (compare Figure 1).

To quantify the dependencies within the network, a conditional probability table (CPT) is allocated to every node. CPTs may be derived from expert opinion or from existing data. Given the probabilities for the parent nodes, the probability of every child node will adapt according to the specific state of a parent node through a process known as "belief updating".

Based on the different probabilities of the nodes in the BN a set of probabilities for the predefined target variable can be calculated. Thus in the network in Figure 1 below, the probability that the macroinvertebrate community will be in each of the four quality classes is shown given an equal chance of six different climates and a set of land use probabilities. The network can then be used to explore the effects of changes in climate and land use on macroinvertebrate quality.

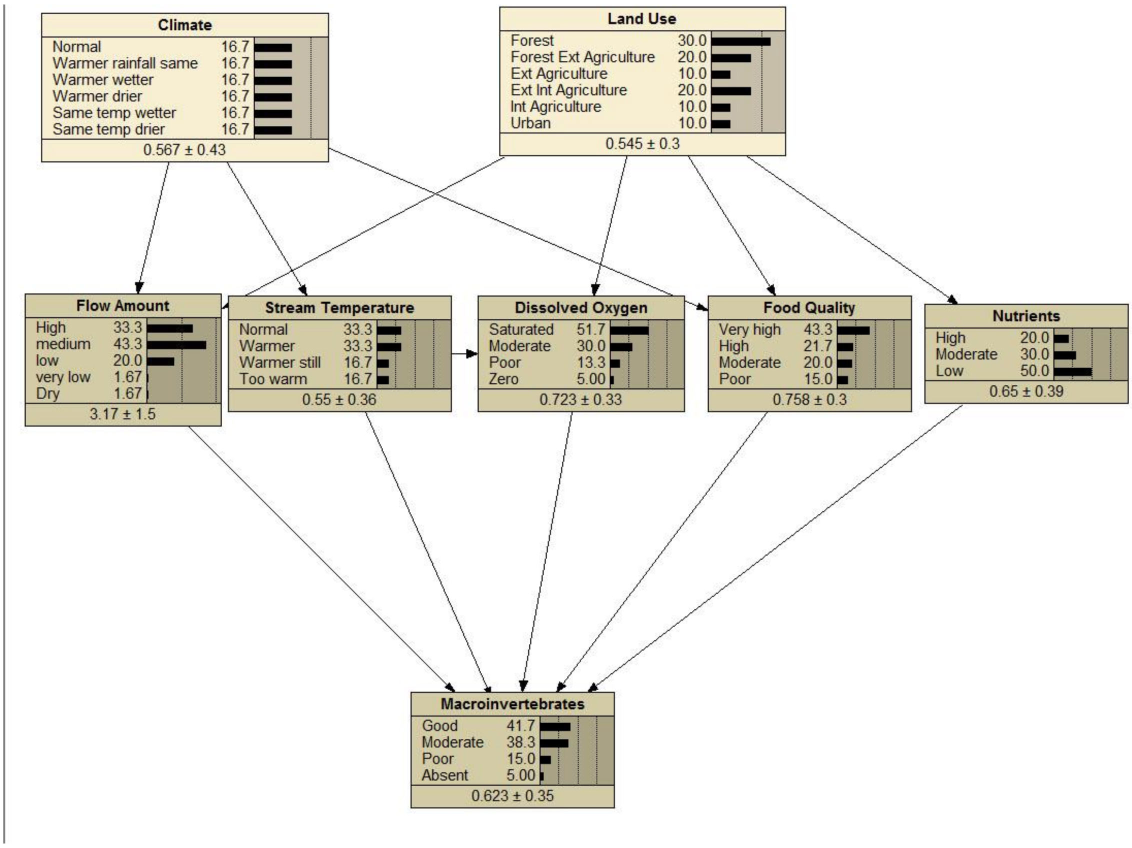


Figure 1: Hypothetical Bayesian Network for the interrelation of environmental variables and the macroinvertebrate community in lowland streams (expert judgements provided by Anna Besse and Piet Verdonschot, Alterra).

To set up a Bayesian network, the following steps need to be accomplished:

1) Define a target (response) variable (or several target variables)
2) Choose appropriate variables affecting the target variable
3) Put the variables in a reasonable order, i.e. a flowchart
4) Define the dependencies between the different variables
5) Define the CPTs for each node
6) Link the resulting network to the output of the abiotic models

Selection of target biological metrics for the study catchments

Initially, partners have been asked to suggest suitable target variables and relationships. The answers to this question are given in the Annex to this document in an unedited form. Based on this a simple overview was generated covering water types, pressures, drivers and organism groups for which data are available (Table 1).

Table 1: Overview of catchment characteristics of the REFRESH study catchments, based on the partner's responses. In the last column the following colour coding is used: Green: Phytoplankton; Yellow: Macrophytes; Purple: Riparian Vegetation, Grey: Macroinvertebrates; Blue: Fish.

Site	**Country**	Water Type	Pressures	Drivers of Change	Data available		Potentially suited explanatory and target variables		Environmental-biotic relationships potentially suited for Bayesian Models
					Organisms	Processes	Abiotic Parameters	Ecological Indicators	
Yläneenjoki Catchment and Lake Pyhäjärvi	Finland	Large shallow lowland lake	Climate change; agricultural intensification	Reducing diffuse loads; abstraction for irrigation; improving rural sewage systems	Phytoplankton biomass; chl a.	None	N and P concentrations and loads; flow.	Phytoplankton biomass (chl a), probably also biomass or proportion of cyanobacteria.	• N/P concentrations → Chl a, cyanobacteria
Vansjø-Hobøl	Norway	Large shallow lake	Climate change; diffuse pollution from agriculture and houses	Nutrients from agriculture and sewage	Phytoplankton, 30 years monthly lake sampling. Also zooplankton, fish, but much less frequent	Primary production, biomass	Total N and P species,	Chlorophyll a, Secchi depth, phytoplankton groups, amount of cyanobacteria	• N/P concentrations → Chl a, cyanobacteria
Tributaries of Lake Simcoe	Canada	Small hard water streams	Climate change; intensive agriculture; urbanisation	Increasing P loadings; atmospheric P deposition	None	None	Nitrate, ammonia, total P, SRP, total suspended solids in the tributaries, and tributary export to the lake.	Use chemical thresholds as surrogate.	• N/A
River Dee	Scotland	Small lowland stream	Climate change; land use change; population increase.	Increasing loadings from agriculture and sewage; decreasing discharge.	Long-term data on fish, esp. salmonids, inverts (10 year record from 16 sites) Coliforms-10 year record	Primary production, respiration (selected sites).	NO_3, P, pH, ANC, temperature, sediment	Salmonids (acidific., sediment.); inverts (acidif., nutrients), Coliforms as indicator for septic tanks etc	• pH, ANC → salmonids and other fish metrics • NO3 → invertebrate metrics • Sediment → Invertebrate metrics

Site	Country	Water Type	Pressures	Drivers of Change	Data available		Potentially suited explanatory and target variables		Environmental-biotic relationships potentially suited for Bayesian Models
					Organisms	Processes	Abiotic Parameters	Ecological Indicators	
Rivers Thames and Kennet	England	Large and medium lowland stream	Climate change; agricultural change; population increase; emission control legislation (for N)	Increasing water abstraction; changes in effluent flows and management; adaptation measures (e.g. new reservoirs)	Diatoms, Suspended Chl a, Macrophyte biomass. Periphyton biomass per macrophyte spp	None	Nitrate, ammonia, total and soluble phosphorus, macrophytes, epiphytes, phytoplankton	Macrophyte, epiphyte and filamentous algal biomass and/or production. Suspended chl-a concentrations	• N / P parameters → macrophyte biomass, algae biomass, Chl a
Vlatava River Catchment	Czech Republic	Stratified reservoirs	Climate change; population distribution change; intensification of agriculture and aquaculture; S and N emission change	Increased P loads from sewage effluent and aquaculture; increased N loads from agriculture; change in precipitation composition leading to increased DOC; temperature and precipitation change leading to water level fluctuations	Phytoplankton, zooplankton, fish, bacteria, macrophytes	Some data on primary production of phytoplankton and bacteria, grazing by flagellates and rotifers.	P, N, DOC, water temperature, runoff	Phytoplankton concentration, macrophyte abundance	• N / P / DOC / temp → macrophyte and phytoplankton biomass
Tordera	Spain	Mediterranean stream, large altitudinal range, variable discharge	Climate change; population increase; agriculture in sections of the catchment	Water abstraction, industrial and urban point source pollution (P, NH_4), diffuse pollution (NO_3)	Invertebrates (>12 years data at family level). Diatoms and fish.	Point measurements of nutrient retention, and stream metabolism	Discharge, nitrate concentration.	Invertebrates	• Discharge → invertebrates • NO3 → invertebrates
Lake Beysehir	Turkey	Large high-altitude lake in semi-arid area	Climate change; land use change; agricultural change	Water abstraction for irrigation; sewage effluent leading to high TP and TN.	No data at present: will have fish, macrophytes, zooplankton and phytoplankton	None	Lake water level, water clarity, nutrients (TP, TN).	Macrophyte abundance, phytoplankton composition	• Water level → macrophytes • Water clarity, N, P → macrophytes, phytoplankton

Site	**Country**	Water Type	Pressures	Drivers of Change	Data available		Potentially suited explanatory and target variables		Environmental-biotic relationships potentially suited for Bayesian Models
					Organisms	Processes	Abiotic Parameters	Ecological Indicators	
					for lake				
River Louros	Greece	Medium Mediterranean stream	Climate change; Land use change: urbanisation, fish farms, recreational areas, agriculture, flood defences; hydroelectric developments.	Water abstraction leading to low flows; reduction of the riparian zone leading to increased water temperature; river channelization and straightening; agriculture leading to increased NH4-N and P.	Aquatic macrophytes and riparian species (3 years).	Trophic / nutrient tolerance of aquatic macrophytes : Primary production (Chl-a)	Nitrate, ammonium, and P concentrations Chl-a , Temperature; Conductivity; Alkalinity.	Macrophyte indices; Riparian integrity indices	• N/P/Chl a → macrophytes • Temp → macropyhtes • N/P → riparian vegetation

Based on this overview, and the more specific answers of the partners giving in the Annex, the general steps towards a Bayesian Network given on page 4 of this document can be simplified as follows:

1. Definition of “candidates” for biological variables to be modelled (i.e. metrics such as Saprobic Indices, traits…) to be used as target variables
2. Screening of environmental variables which are modelled in the catchment
3. Construction of a Bayesian Network visualizing relationships between environmental and biological (target) variables
4. Quantification of the individual linkages between environmental and biological variables, ideally with data from catchments of the same “river type” or “lake type”.

For some catchments the model concepts are already advanced:

- For Lake Pyhäjärvi (Finland) relationships between nutrients and phytoplankton composition is modeled in a mechanistic way
- For Lake Vansjø-Hobøl (Norway) Bayesian Networks have already been constructed
- For the wetlands of the Louros catchment (Greece) a detailed concept has been developed (compare Annex)

Thus, steps 1 to 4 need to be defined primarily for the following catchments:

- River Dee (Scotland)
- Rivers Thames and Kennet (England)
- Vlatava river catchment (Czech Republic)
- Tordera (Spain)

The selected indicator groups are heterogeneous, reflecting data availability in the catchments. In total, five different organism groups are covered (compare Table 1):

- Phytoplankton / Chl a / algae: 5 cases
- Macrophytes: 6 cases
- Riparian vegetation: 1 case
- Invertebrates: 4 cases
- Fish: 1 case

The response of these organism groups, however, needs to be measured by one or several indices (metrics). These metrics should meet the following criteria:

- They need to be compliant to the Water Framework Directive (WFD), ideally they should be the “standard” methods used in the country for the WFD monitoring. A database of all methods used for WFD implementation in the individual European countries can be found on http://www.wiser.eu/programme-and-results/data-and-guidelines/method-database/.

- Alternatively, "intercalibration common metrics" could be used, which are accessible through the website mentioned above. These are more generally applicable metrics which have been used to compare assessment systems of different European countries.
- In any case, the metrics should meet the taxonomic resolution usually applied in the country for the organism group in question, e.g. family level for invertebrates in the UK or species level for invertebrates in the Czech Republic.

Consequently, the following metrics are suggested as surrogates for the biota of the study catchments (Table 2).

Table 2: Metrics suggested as target variables for the Bayesian Networks to be developed for the catchments Dee, Thames/Kennet, Vlatava and Tordera

Catchment	Dee (Scotland)
Stressor	Climate change; land use change; population increase; acidification
Organism group(s)	Fish, invertebrates
Metric candidates	- Fish: No metric-based standard methods available in the UK; simple metrics such as abundance of target species (trout, salmon) might be suited - Invertebrates: ASPT (pollution), LIFE-Index (flow, sediment)
Environmental variables	NO_3, P, pH, ANC, temperature, sediment
Possible data sources for conditional probability tables	Data sources from River Dee (long term monitoring of fish and invertebrates) Routine monitoring programmes from neighbouring catchments (?)

Catchment	Thames / Kennet (England)
Stressor	Climate change; agricultural change; population increase; emission control legislation (for N)
Organism group(s)	Diatoms, Suspended Chl a, Macrophyte biomass, Periphyton biomass per macrophyte spp (modelled); invertebrates (data should be available from routine monitoring)
Metric candidates	- Diatoms: WFD River Diatom method or Trophic Diatom Index version 3 Method - Macrophytes: River Macrophyte Nutrient Index; River Macrophyte Hydraulic Index, Functional Group Diversity, Number of Taxa, Filamentous Algal Cover (all part of UK Ecological Classification of Rivers using Macrophytes) - Invertebrates: ASPT (pollution), LIFE-Index (flow, sediment)
Environmental variables	Nitrate, ammonia, total and soluble phosphorus
Possible data sources for conditional probability tables	Routine monitoring data from England and Wales, intercalibration datasets

Catchment	Vlatava (Czech Republic)
Stressor	Climate change; population distribution change; intensification of agriculture and aquaculture; S and N emission change
Organism group(s)	Phytoplankton, zooplankton, fish, bacteria, macrophytes (reservoirs?); invertebrates, diatoms, fish (data should be available from routine monitoring)
Metric candidates	- Invertebrates: Czech system for ecological status assessment of rivers using benthic macroinvertebrates [Systém pro hodnocení ekologického stavu toku podle makrozoobentosu]; about 15 metrics included, amongst others the Czech Saprobic Index - Macrophytes: no official system recorded (?) - Phytobenthos (diatoms): Assessment system for rivers using phytobenthos [Hodnocení tekoucích vod podle fytobentosu]; using the Czech saprobic-trophic index - Fish: Czech national method of the river ecological status classification according to the fish biocoenosis [Ceský index hodnocení ekologické kvality toku pomocí rybích spolecenstev]; metrics used: Presence of typical species; Overall abundance; Relative abundance of rheophilic species; Relative abundance of eurytopic species
Environmental variables	P, N, DOC, water temperature, runoff
Possible data sources for conditional probability tables	Czech routine monitoring data (?); STAR data

Catchment	Tordera (Spain)
Stressor	Climate change; population increase; agriculture in sections of the catchment
Organism group(s)	Invertebrates (>12 years data at family level), diatoms and fish
Metric candidates	- Invertebrates: Iberian BMWP - Diatoms: Pollution Sensitivity Index - Fish: Index of Biotic Integrity using fish as indicators of the Ecological Status of Catalonian Rivers
Environmental variables	Discharge, nitrate concentration
Possible data sources for conditional probability tables	Routine monitoring programmes?

Next steps towards the development of Bayesian Networks

Screening of environmental variables which are modelled in the catchment

Only a limited number of environmental variables are modelled in the individual catchments, usually on a very high temporal and spatial resolution. An overview of the variables is given in Table 1. It might be necessary to refine this also giving the spatial and temporal resolution achieved in the models. Further, this list might be supplemented by additional environmental variables which are not modelled but impact the organism group.

Construction of a Bayesian Network

Bayesian Networks need to be constructed for each individual catchment, visualizing the linkages between the individual environmental variables and the biological metrics. A simple example is given in Figure 1.

Quantification of the linkages between environmental and biological variables

Each of the linkages of the Bayesian Network needs to be subject of a conditional probability table. I suggest that the classes of the biological metric are given in the WFD nomenclature, either in five classes (high, good, moderate, poor, bad) or in two (high/good, moderate-bad). Also, the values of the environmental variables need to be broken down into classes. This is a more individual decision, based on the on range of values observed in the catchment.

The conditional probability table will give the number of observations for each combination of class of an environmental (explanatory) variable and the biological (response) variable. The number of observations is to be taken from an independent (ideally large) data source from a comparable stream / lake type, e.g. from standard WFD monitoring programmes.

References cited

Charniak E. (1991) Bayesian networks without tears. AI Magazine, 12, 50–63.

Stewart-Koster B., Bunn S.E., Mackay S.J., Poff N.L., Naiman R.J., Lake P.S. (2010) The use of Bayesian networks to guide investments in flow and catchment restoration for impaired river ecosystems. Freshwater Biology, 55, 243–260.

Annex: Partner responses on modelling approaches to link environmental and biotic data in the study catchments

Jannicke Moe (NIVA)

We propose a procedure which builds upon a previous modelling work with Lake Vansjø (in collaboration with Bioforsk). The modelling approach in this project was Bayesian probability network, which we could say integrates a set of knowledge rules using conditional probability tables (CPT). For modelling of ecological indicators in REFRESH we don't propose using the network modelling method itself, but we could use similar steps for setting up knowledge rules.

A brief description of the procedure:

1) Select a response variable (e.g. %cyanobacteria) and one or more pressure variables (e.g. total P and temperature).

2) Discretise the variables into intervals (see more information below), to produce a conditional probability table.

3) For each table cell (each combination of pressure(s) and response intervals), calculate the proportion of observations that falls into this cell (see Figure 1 in attached document).

Such a conditional probability table could in itself function as a kind of knowledge rule. The step from pressure value(s) to response value can then be modelled in a simple way, e.g. by a set of "if -else" statements.

For cases where data are not quantitative, the variable can instead be yes/no or other kind of categories (for both pressure and/or response variables). This approach is very general, and should work for most types of data and case studies.

In addition, conditional probability tables can be used in more flexible ways in a Bayesian network (BN) model, which we will probably do at NIVA (as we discussed briefly in Antalya). In the graphical version of a BN, the conditional probability table is represented by an arrow between two or more nodes (variables). One of the strengths of a BN is that it can also be run "backwards": from output variable to input variable. For example, we can set the %cyanobacteria (output) to a certain interval of interest, and get the corresponding probability distribution of the total phosphorus concentration intervals (input).

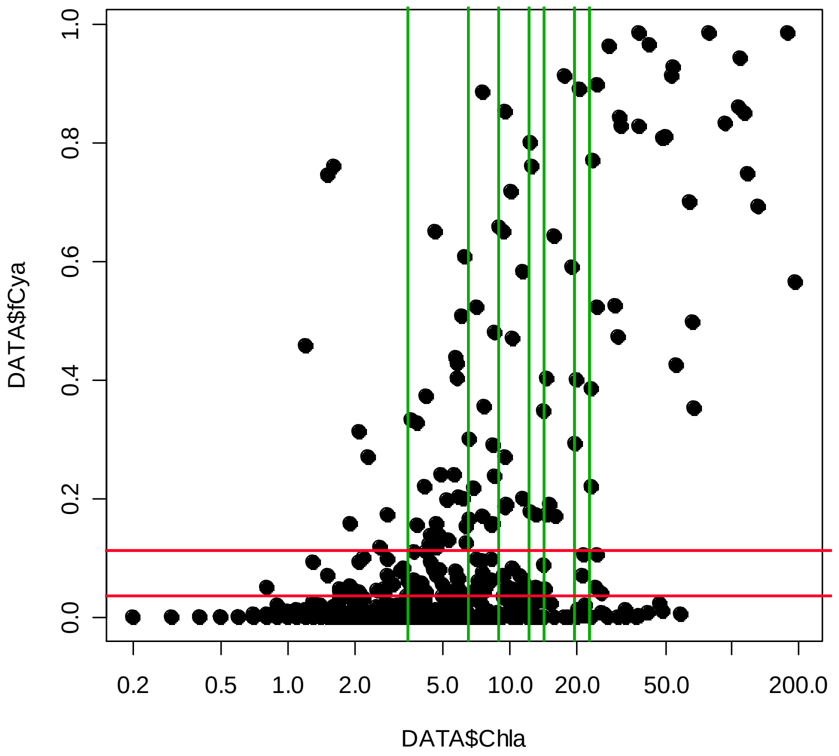
Discretisation of variables:

Different criteria can be used for setting the intervals of the variables. We have used the statistical method regression tree (see Figure 2), which can optimise the intervals of the pressure variable in such a way that the variation of the response variable is low within the intervals and high among the intervals. This regression method is not dependent on specific data distributions, and should therefore be more robust than ordinary regression methods.

Other alternatives can be e.g. to select regular intervals (e.g. TotP 0 - 5 - 10 - 15 etc.), or to select intervals which have some particular interest (e.g. values of %cyanobacteria corresponding to WFD status class boundaries).

In general, a higher number of intervals require more data for a reliable"parametrisation" of the conditional probability table.

Water samples for the analysis of phytoplankton are taken as integrated water samples (0-4 m depth) with a Ramberg water sampler (NIVA, Oslo). The samples are collected in 100 mL glass bottles and fixed immediately with Lugol's iodine solution.



%cyanobact, Method 1

Lake Chl. [...	0 - 3	3 - 6	6 - 9	9 - 12	12 - 15	15 - 18	18 - 21	21 - inf
0-5%	0.854167	0.77193	0.606061	0.666667	0.6	0.6	0.5	0.346154
5-10%	0.083333	0.122807	0.181818	0.148148	0.2	0	0	0.076923
10-100%	0.0625	0.105263	0.212121	0.185185	0.2	0.4	0.5	0.576923

Figure 1: Conditional probability table based on the number of observations in each combination of intervals of %cyanobacteria (response variable) and Chlorophyll a (pressure variable).

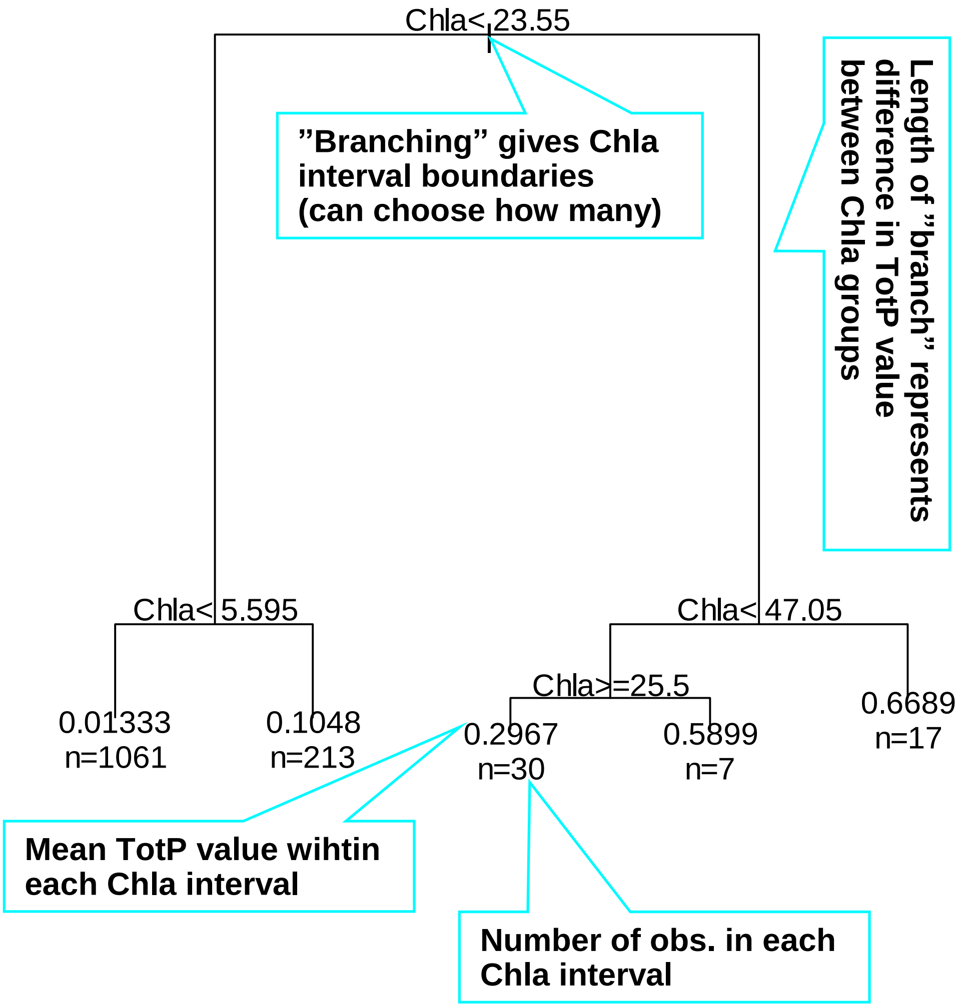


Figure 2: Regression tree for discretisation of the variables chlorophyll a (response) and total phosphorus (pressure).

Ahti Lepistö, Olli Malve, Niina Kotamäki and Marko Järvinen (SYKE)

'Knowledge rules' are functions that would allow us to link chemical concentrations to ecological indicators. A key question is how these functions change in changing climate. Lake Pyhäjärvi is the largest lake in South-Western Finland. It is shallow and mesotrophic with exceptionally high fish productivity (Sarvala et al., 1998), which makes it an important lake also for commercial fishing. Increased eutrophication of the Lake Pyhäjärvi has been a major concern since the late 1980s as cyanobacterial blooms have become more frequent. Chlorophyll-a concentration indicating phytoplankton biomass has been increasing in 1980-2005: from the level of 4-6 µg l^{-1} to 6-10 µg l^{-1} , together with the increase in totP concentrations and decrease in Secchi depth (Fig. 1) (Ventelä et al., 2007). Two approaches to link nutrient concentrations to water ecology (principal indicator Chl-*a*) are briefly discussed 1) Empirical approach and 2) LLR model approach.

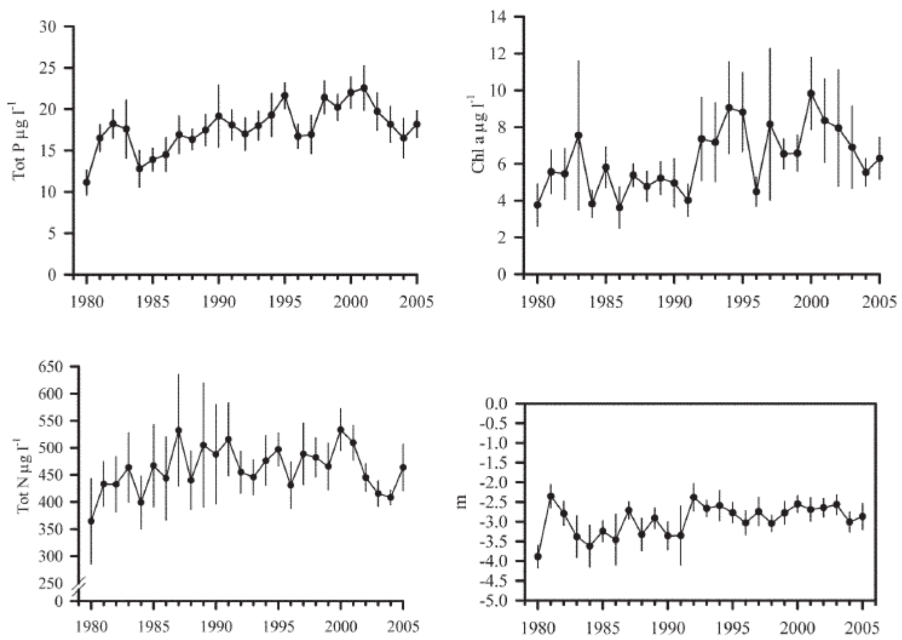


Fig. 1. Mean total P, total N and Chl-a concentrations and Secchi depth in Pyhäjärvi during the open water season (May-Oct) in 1980-2005. Vertical lines denote 95% confidence limits (Ventelä et al., 2007).

Empirical, statistical approach

The tentative conceptual chart (Fig. 2) illustrates possible interactions between phytoplankton blooms and environmental variables. The focus in this indicator analysis of Pyhäjärvi will be on the right-bottom corner of the chart (blue colour), i.e. the functions that would allow us to link nitrogen and phosphorus concentrations to the ecological indicators. Chlorophyll-a concentration is among the most important indicators of eutrophication in the lake, as also demonstrated by the linear relationship between the mean summer total-P and chl-a (Fig. 3). The relatively low (r^2=0.47) regression between the variables however indirectly indicates the role of internal nutrient loading in the lake, which gets more evidence from the lake float highfrequency monitoring results. These results indicate increased cyanobacterial abundance after the mixing events in the shallow lake parts of the lake (resuspension). Phytoplankton blooms typically occur in the lake during late summer (from mid-August to September) and their occurrence is highly variable between the years. Lake Säkylän Pyhäjärvi has been a target lake for many means of lake restoration, including the food-web manipulation. These restoration efforts, together with high variability in external and internal loading, increase year-to-year variability between total-P and chl-a.

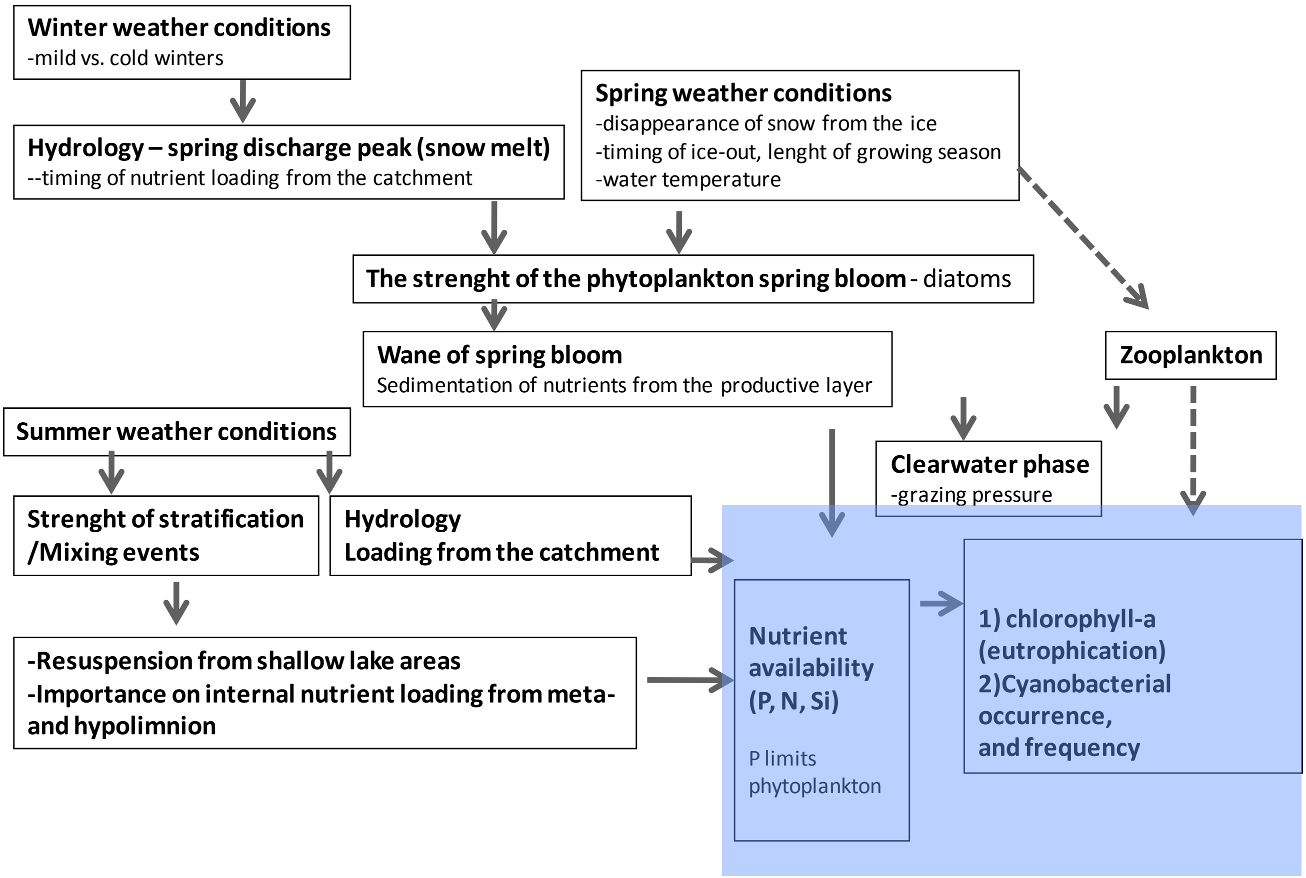


Fig. 2. A conceptual model illustrating potential interactions between the environmental variables and phytoplankton occurrence in Säkylän Pyhäjärvi.

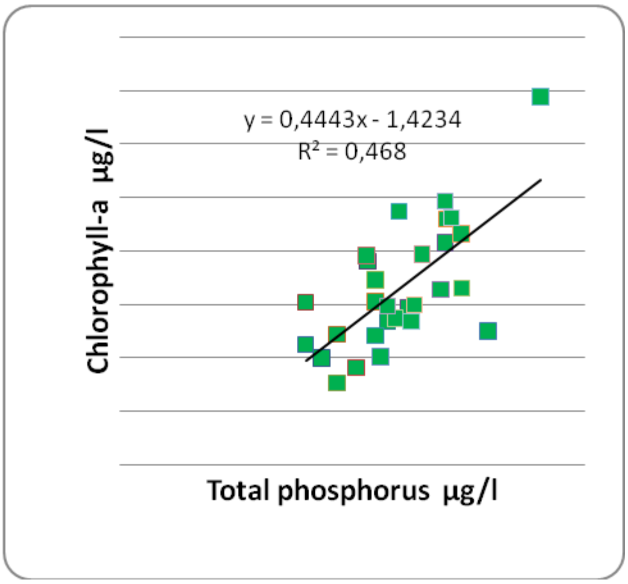


Fig. 3. Linear relationship between the mean summer (Jun-Sept) lake total P concentration and chl-a concentration in Lake Pyhäjärvi (1977-2008).

LLR – Steady state probabilistic box model

The Lake Load Response (LLR) internet tool has been developed at SYKE to ease the use of models in making predictions about the effects of nutrient loading into a lake. LLR consists of three steady state models:

- Chapra's (1975) model for retention of total phosphorus and nitrogen including probabilistic estimates for model parameters and model predictions. It contains also different options for settling rate calculations,
- the hierarchical, linear regression model for chlorophyll-*a* (Malve & Qian 2006), and
- the logistic regression model for phytoplankton biomass

With Chapra's retention model it is possible to estimate the in-lake nutrient concentrations as a function of incoming load to the lake and water outflow. 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). From the relationship between nutrient and chlorophyll-a concentration in the lake, it is possible to determine the relation between nutrient loading and chlorophyll-a concentration. This leads to predictions about the target load with which a good or high ecological status of the lake according to chlorophyll-a concentration can be 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 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, chlorophyll-a prediction is based almost solely on the data from the study lake itself, when there is plenty of data available. But if there is a very high variability for any reason, the predictions are based on the chlorophyll-a response to nutrient concentrations within the lakes of same type. 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.

Preliminary LLR model runs were conducted. First, the LLR model was applied to predictions of total phosphorus and nitrogen concentrations as functions of loads (Figures 4 and 5). Second, the model was applied to study the effect of nutrient load to chlorophyll- *a* in Lake Pyhäjärvi, and to estimate target nutrient loads given the lake type specific Good/Moderate (and High/Good) boundaries of TP, TN and chlorophyll a (Fig 6). Different probabilities are also shown (50% / 90%), as well as Chl-a estimate with the present load.

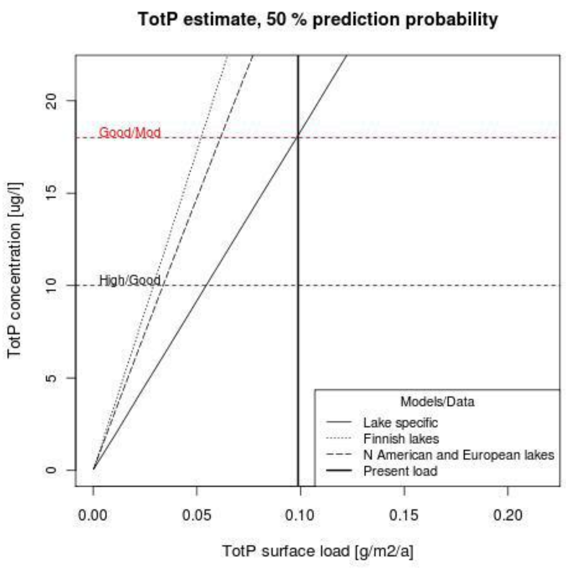


Fig. 4 Estimate for total phosphorus concentration (μg/l) in Lake Pyhäjärvi as a function of loading (g/m2/a). Red horizontal dash line is the limit for good water quality according to WFD classification in Finland. The vertical indicates the present loading.

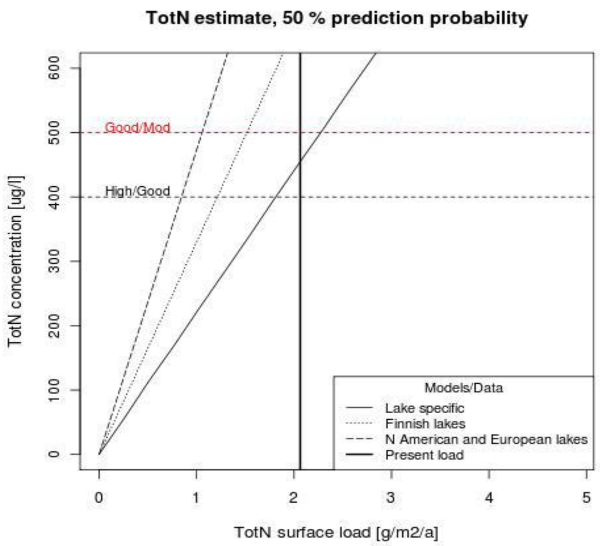


Fig. 5 Estimate for nitrogen concentration (µg/l) in Lake Pyhäjärvi as a function of loading (g/m2/a). Red horizontal dash line is the limit for good water quality according to WFD classification in Finland. The vertical indicates the present loading.

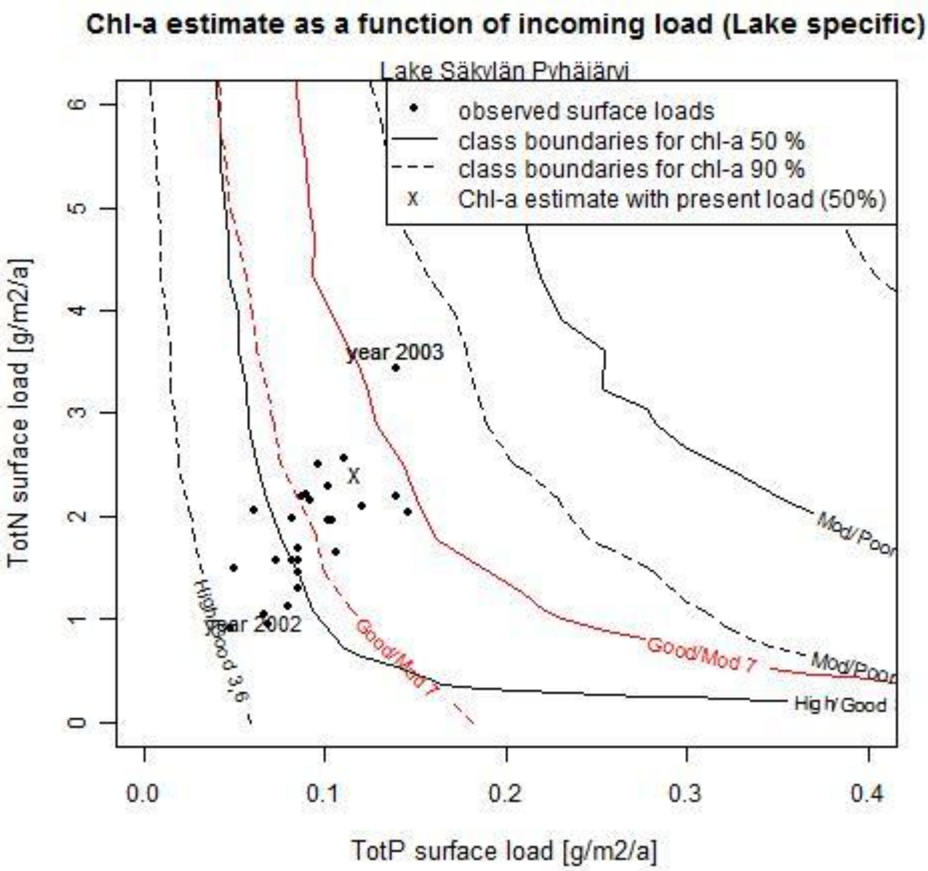


Fig. 6 LLR-model estimate for chlorophyll-a concentration in Lake Pyhäjärvi as a function of phosphorus and nitrogen loading (g/m2/a) to the lake. The red curve shows totP / totN combinations with which the Chl- a concentration will stay at good water quality

(Good/Moderate, chl-a 7 µg l^{-1}) with 50 % probability. The High/Good and Moderate/Poor -limits are also shown, as well as different probabilities (50% and 90%). x shows the Chl-a estimate with present load, and extreme years of 2002 and 2003 are shown separately.

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 Resources Research 11: 1033–1034.

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

Lamon, E.C., Malve, O. and 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.

Sarvala, J., Helminen, H. and Auvinen, H. 1998. Portrait of a flourishing freshwater fishery: Pyhäjärvi, a lake in SW-Finland. Boreal Environment Research 3: 329–345.

Ventelä, A.-M., Tarvainen, M., Helminen, H. and Sarvala, J. 2007. Long-term management of Pyhäjärvi (southwest Finland): eutrophication, restoration – recovery? Lake and Reservoir Management 23: 428–438.

Eva Papastergiadou (University of Patras, Greece)

Green text is from the original request for suggested indicators from Daniel Herring and Richard Skeffington

1. Answer the questions in the Indicators Document i.e. for the biotic data available

Are these data qualitative or quantitative?

Aquatic macrophytes and riparian plants. Quantitative 5-degree scale (5: highly abundant, 0: absent).

Data collected seasonally from 18 sampling sites starting from spring 2005 ending on November of 2007. For spring 2006 two samplings were conducted. Sum: 3 years X 9 seasons X 18 sites = 486

Taxonomic resolution (family, genus, species)

Species level (macrophytes and riparian plants)

- Sampling method applied

Representative river reaches were selected in order to conduct the plant surveys. Sites of varying degrees of apparent human impacts on the river's main stem and tributaries were surveyed. The assessment of river started from the springs (Terovo) and extended to the estuaries (Amvrakikos gulf). The survey area was divided in 3 zones, a) the wetted part of the channel, b) the marginal-active channel and c) the riparian woodland plot. A homogenous reaches of 100m river length were sampled taking account hydrological & morphological characteristics and all species found in the three zones were included in the survey plan.

Aquatic macrophytes. [a ''*Macrophyte''* is defined as 'any plant observable with the naked eye, including larger algae, bryophytes, and vascular plants']. Following a shoreline walk in each sampling point, and in some cases with boat, the abundance of aquatic macrophytes were recorded at the site based on semi-quantitative DAFOR scale where: D=Dominant, A=Abundant, F=Frequent, O=Occasional και R=Rare (Palmer, 1992).

Riparian species

For the riparian vegetation survey we focused on the streamside woodland area within the wider riparian corridor. The assessed riparian plot length was standardized at 100m according to the widely Mediterranean assessment method QBR (Munne *et al*., 2003) but the width varies with the extent of the riparian woodland. The width of the plots always begins at the end of the active channel and usually ends near the lateral extent of the riparian woodland. Within each plot the percentage cover of all plants was estimated. The percentage of each plant species, was assessed, using a 5-point scale, by a thoroughly walking the entire plot, sometimes criss-crossing twice or more times through the woodland to locate all species. In each site environmental characteristics where recorded considering: i) *Slope* (%), Strahler's order (Strah_Ord), and Watershead area (Wat_Area) were estimated using Geographical Information

System (GIS), ii) *Distance from nearest spring* (m) (Spr_Dist), iii) *Distance from sea* (m) (Sea_Dist) were measured using Topographic maps 1:50 000. Information about iv) mean *annual precipitation (*P_annual) and v) main *air temperature* (T_annual) and *Discharge* were taken from the nearest weather stations.

This would help in searching for similar data from comparable catchments

2. Have you got data which could be used to set up a knowledge rule for your particular catchment?

We tested the following candidate metrics:

Macrophyte metrics

i) Abundance/composition

- Total number of species
- Total number of families
- Total number of genus
- Number of Hygrophilous sp.
- Number of Helophyte sp.
- Number of Red Algae
- Number of Green Algae
- Number of Bryophytes sp.
- Number of Pteridophytes sp.
- Number of Hydrophytes sp.
- Number of parvopot sp.
- Number of ceratophilids sp.
- Number of myriophyllids sp.
- Number of magnopot sp.
- Number of natopot sp.
- Number of hydroch sp.
- Number of lemnids sp.

ii) Diversity and Richness

- Shannon-Wiener diversity index
- Simpson's Index

- Evenness
- Species Richness

i) Trophic indicators

  - Number of Nitrophyllus sp.
  - IBMR
  - MTR

Riparian integrity indices

- Riparian woodland [%]
- QBR
- Total Riparian width (m)
- Lack of natural wooded vegetation [yes/no]
- *Platanus orientalis* cover
- Riparian vegetation zone continuity with terrestrial woodland
- Number of woody riparian sp.

Sensitivity/Tolerance

- *Arundo donax* cover [%]
- *Rubus spp* cover [%]
- Number of exotic sp.
- Number of ruderal sp.
- Number of weedy sp.

3. If so, could you attempt to produce a draft knowledge rule? This would be a draft for comment rather than the finalised version.

- N/ P/ Chl a → macrophytes
- Temp → macropyhtes
- N/ P → riparian vegetation

Draft knowledge rule

Hydrological and physical disturbances are particularly important in Mediterranean-type Rivers, due to the water constraints and the long-term agricultural use of river surroundings.

- Species predictive responses to disturbance e.g. use of MAC and MACPACS predictive models

(Francisca C. Aguiar, Maria João Feio, Maria Teresa Ferreira: **Choosing the best method for stream bio assessment using macrophyte communities: Indices and predictive models** Ecological Indicators 2010.

The predictive models developed in the present study used the same functional species groups, had similar reference vegetation types. Compared to MAC, MACPACS has the advantage of incorporating fewer and more easily available predictive variables, such as altitude and yearly average rainfall. MACPACS also classifies sites into five bands of increasing deviation from reference conditions, which can be used as proxies of WFD ecological quality classes. The use of species presence/absence means that less field- sampling work is needed compared to any of the other studied approaches. However, MACPACS was not as effective as MAC at predicting degradation. The difference between the predictive models may be due to the fact that while **MAC** considers the abundance for each species, **MACPACS** only considers their presence or absence at the site. We therefore think that abundance was the key factor in the efficiency of the MAC model. The incorporation of abundance data in vegetation predictive models appears to be particularly important to the detection of high levels of degradation. For instance, if a site has an excessive growth of certain species such as alien macrophytes, species occurrence (i.e. using MACPACS) will be insufficient for a suitable assessment status.

- The macrophyte regression model

(Iris Baart, Christine Gschφpf, Alfred Paul Blaschke, Stefan Preiner, Thomas Hein. **Prediction of potential macrophyte development in response to restoration measures in an urban riverine wetland.** Aquatic Botany 2010.)

The present study examined the species diversity and abundance of macrophytes in relation to hydrological parameters. A macrophyte regression model was developed based on the output of a 2D hydraulic model for different wetland management options The macrophyte model was developed using stepwise multiple linear regressions between macrophyte variables and environmental and hydrological parameters with the software SPSS 15. The stepwise procedure was preferred as it combines forward and backward methods iteratively (Bóhl and Zφfel, 2005). Potential autocorrelations were first surveyed with PCA. The final regression models were integrated into the GIS framework (ArcGIS, ESRI) to create digital maps of the distribution of the macrophyte indicators of abundance (as total cover of macrophytes in percentage) and species richness per survey section.)

- Conceptual model

J. M. O'HARE M. T. O'HARE A. M. GURNELL M. J. DUNBAR, P. M. SCARLETT and C. LAIZE. **Physical constraints on the distribution of macrophytes linked with flow and sediment dynamics in British rivers.** River. Res. Applic. (2010)

A conceptual model was produced which indicates ranges of SSP which may determine the significance of aquatic macrophytes in channel engineering processes. This model could contribute to predicting the potential for macrophyte growth within a given reach thus indicating its capacity for self-restoration or the likelihood of weed problems. Specific stream power (SSP) was calculated to represent hydrological disturbance and a median bed calibre index and percentage sand and finer sediment were used to characterize substrate size, since stream energy and sediment properties are two major physical controls on aquatic vegetation.

- A multiple regression model

Derek R. Shiels. **Implementing landscape indices to predict stream water quality in an agricultural setting: An assessment of the Lake and River Enhancement (LARE) protocol in the Mississinewa River watershed, East-Central Indiana.** Ecological Indicators 2010.

A multiple regression model was run for each of the dependant response variables: TP, E.COLI, NO3 or EPT/C. For three of these models, a total sample size of thirty sub-watersheds was available; the NO3 model only had 25 sample sub-watersheds due to NO3 data in only three of the four LARE studies. The sample sites were compiled from studies completed by four separate entities in 2001, 2003, 2005 and 2007. Therefore, there was need to account for the variation (sampling methods, season, flow rates, etc.) that existed among the four studies. This was accomplished by adding extra predictor variables labeled as "procedure" (Px) variables to the models. Hence, in this study, in addition to six landscape indices, three "procedure" (P1, P2, and P3) variables composed the total 9 (8 for NO3) predictor variables.

- Riparian vegetation-flow response guilds

DAVID M. MERRITT, MICHAEL L. SCOTT, N. LEROY POFF, GREGOR T. AUBLE & DAVID A. LYTLE. **Theory, methods and tools for determining environmental flows for riparian vegetation: riparian vegetation-flow response guilds.** Freshwater Biology (2010) 55, 206–225

The development of riparian vegetation-flow response guilds offers a framework for transferring information from rivers where flow standards have been developed to maintain desirable vegetation attributes, to rivers with little or no existing information. We propose to organise riparian plants into non-phylogenetic groupings of species with shared traits that are related to components of hydrologic regime: life history, reproductive strategy, morphology, adaptations to fluvial disturbance and adaptations to water availability. Plants from any river or region may be grouped into these guilds and related to hydrologic attributes of a specific class of river using probabilistic response curves. Probabilistic models based on riparian response guilds enable prediction of the likelihood of change in each of the response guilds given projected changes in flow, and facilitate examination of trade-offs and risks associated with various flow management strategies. Riparian response guilds can be decomposed to the species level for individual projects or used to develop flow management guidelines for regional water management plans.

- Conceptual model of a stream/river ecosystem and its riparian elements

R. Brooks M. McKenney-Easterling M. Brinson R. Rheinhardt K. Havens D. O'Brien J. Bishop J. Rubbo B. Armstrong J. Hite. **A Stream–Wetland–Riparian (SWR) index for assessing condition of aquatic ecosystems in small watersheds along the Atlantic slope of the eastern U.S.** Environ Monit Assess (2009) 150:101–117.

We followed the lead of Brooks et al. (1998, 2006a) in considering streams, wetlands, and riparian areas as definable landscape units that support characteristic water-dependent biota (i.e., stream- and wetland-dependent species of invertebrates, vertebrates, and vascular plants) and that respond predictably to a set of anthropogenic stressors. We believe that such an approach will assist those concerned with their protection, conservation, and management. The interactive relationships among the stream, wetland, riparian and upland components of watersheds for different stream orders are illustrated in Fig. 2a–c. A key feature of these illustrations is the relative contribution to the functioning of these systems by upstream portions of the watershed, versus immediately adjacent or lateral components. Our approach to assessing condition, presented here as a first step toward understanding the linkages among aquatic components, involves three levels of effort that increase in detail and diagnostic reliability as data collection shifts from remote-sensing to intensive sampling on the ground (Brooks et al. 2004, 2006b). A Level 1 or Landscape Assessment can be accomplished in the office using only remote-sensing data and geographic information systems (GIS). A Level 2 or Rapid Assessment builds upon the findings at Level 1 by adding rapidly implemented ground reconnaissance at the site level. A Level 3 or Intensive Assessment typically requires more intensive data collection, involving HGM functional models (Smith et al. 1995), IBIs (Karr and Chu 1999), or other labor-intensive methods. As anticipated, the degree of confidence in the data used and the reliability of decisions made based on those data increases with greater amounts of effort. However, the spatial coverage of Level 3 data will typically be lower, given the greater level of data collection effort required.

- The models: ‘‘VSG’’ model, RAIS model (Riparian Aquatic Interaction Simulator)

(Paul L. Ringold, John Van Sickle, Mike Bollman, Jeff Welty, Jerry Barker. **Riparian forest indicators of potential future stream condition**. Ecological Indicators 2009.)

The ‘‘VSG’’ model (Van Sickle and Gregory, 1990) estimates the delivery of tree boles to the stream from sources of tree fall such as windthrow, decomposition, and stream bank erosion. Trees were assumed to fall independently of each other; multiple-tree input events such as debris flows and landslides were not considered. Estimates of wood input (bole number and volume) were modeled as functions of a riparian stand's species mix, effective tree height, stem density, and of the assumed probabilities that trees fall (the assumption is made that all trees fall, thus simulating an episode of complete mortality of the current stand) and upon falling land in a userspecified direction.

The RAIS model (Riparian Aquatic Interaction Simulator) (Welty et al., 2002) was chosen to evaluate the implications of stand growth and future mortality. This model combines the mechanisms of tree fall incorporated in the VSG model with a tree growth model, ORGANON (Hann et al., 1995). This simulator describes the growth of the trees and simulates rates of mortality over time, particularly as a function of stand density. Thus, while the two models appear to predict the same thing, they are predictions that reflect very different processes and are not directly or simply comparable to one another. The VSG model is used to describe the potential stream wood contribution given complete mortality of a current stand, while RAIS describes the contribution of stream wood that would result from chronic mortality over a specified period as the trees in the stand grow. Both models have been widely cited and used. The RAIS model, for example, is the foundation of much of the analysis of riparian management in Washington state (Washington Department of Natural Resources, 2001).

- Change of riparian zones assessed by multi-temporal AVHHR datasets.

Eva Ivits, Michael Cherlet, Wolfgang Mehl, Stefan Sommer. **Estimating the ecological status and change of riparian zones in Andalusia assessed by multi-temporal AVHHR datasets.** Ecological Indicators 2009.

We suggest that extensification and thus better ecological status of the riparian zone can be partly approximated by the amount of vegetation permanently present on the area. For this the so-called permanent vegetation fraction was derived from a multi-temporal advanced very high-resolution radiometer (AVHRR) dataset and was used (1) to classify the ecological status of the riparian zone into two classes, favourable and unfavourable, and (2) to assess the effect of agricultural practices on these areas.

- Modelling environmental indicators of stand density, tree growth and health

Patricia Maria Rodriguez-Gonzaleza,, John Christopher Stella, Filipe Campeloc, Maria Teresa Ferreiraa, Antonio Albuquerque. Subsidy or stress? **Tree structure and growth in wetland forests along a hydrological gradient in Southern Europe.** Forest Ecology and Management 259 (2010) 2015–2025

In order to test the effects of all environmental variables on stand structure and tree growth and health, we consolidated the soil variables using a Principal Components Analysis (PCA) using CANOCO4.5 (ter Braak and ˇSmilauer, 2002). All environmental variables shown in Table 1 were used in the PCA except redox potential (because of several missing values), organic phosphorus (which was highly correlated with extractable P) and nitrogen and carbon contents (which were substituted by C:N ratio). We used the first three principal components of the PCA as explanatory factors to model four vegetative indicators of aboveground productivity and tree health: (1) tree density, in trees/ha; (2) shrubbiness growth form, measured as the plot-based average of live stems per tree; (3) stem mortality, measured as the proportion of dead basal area per tree; and (4) recent average residual growth rate (1999–2003). Each of the two species was modelled separately. Linear mixed-effects models were implemented in R 2.5.1. (lme function, R Development Core Team, 2007; Pinheiro and Bates, 2000) using forest stand as a random effect to account for the nesting of plots within stands. For model selection, we took an information-theoretic approach, using Akaike information criterion (AIC) to compare a suite of competing models (Burnham and Anderson, 2002). The best models were selected using the lowest AICc values in a candidate set, and alternate models were assessed using differences from the minimum AIC (_i) and associated Akaike weights (wi). The best model in each candidate set has the lowest _i and highest weight. For each response factor tested (tree density, shrubbiness, growth rate and mortality), we constructed a set of candidate models that included all additive combinations of the first three principal components of the PCA (eight models total).