Application of the Particle Filter to COHERENS V2 model of Lake Säkylän Pyhäjärvi

Akiko Mano* and Takayuki Shuku†

*Ochanomizu University, 2-1-1, Ohtsuka, Bunkyo-ku, Tokyo 112-8610, Japan,
mano.akiko@ocha.ac.jp
† Graduate School of Environmental and Life Science, Okayama University, Japan
shuku@cc.okayama-u.ac.jp

Abstract:

The particle filter (PF, Gordon et al, 1993; Kitagawa, 1996) was applied to determine the optimal parameters for the sediment transport for the lake Säkylän Pyhäjärvi model with COHERENS Version 2. After having obtained the optimal parameters, time series of the TSM (Total Suspended Matters) concentrations at the automatic station were compared among the model result of (1) before- (2) after applying PF and (3) the monitored observation data. The model result with using the optimal parameters after applying PF was well-reproduced the observation data. The application of PF can be considered as a quite usable way to improve the model reproducibility.

1. Introduction

Our research was focused on suspended solid simulation in Lake Säkylän Pyhäjärvi (Fig. 1), the largest lake in southwestern Finland, with the COHERENS V2 model (COHERENS version 2(Luyten, 2011)). The aim of this study is to apply PF to the model to determine the optimal parameters related to sediment transport to improve the model reproducibility.
The lake has two inlets from Ylaneenjoki and Pyhäjoki rivers and one outlet from Eurajoki river. These inflow and outflow river water and meteorological conditions such as wind are considered to be the main driving forces of water circulation in the lake.

2. Data and materials

To put it plainly, the model itself and all data such as the boundary conditions, the initial conditions and the observation data were based on ones used in satellite data assimilation (Mano, 2013) as described below.

The boundary conditions at river mouths were used the following data. Time series of river discharge and TSM loading at each river mouth were extracted from Hertta-database of Finnish

Environmental Administration. Those of temperature were done from VEMALA, the water quality component of the Watershed Simulation and Forecasting System (Vehviläinen B, et al., 2005) of the Finnish Environment Institute. Meteorological data was obtained from Finnish Meteorological Institute which contains wind speed and direction, air temperature, humidity, cloud coverage and air pressure. The initial condition for the TSM concentration on the lake surface was converted from Turbidity of the satellite data on June 26 in 2009 constructed in Finnish Environment Institute. The observation data was monitored on the surface of the lake at the automatic station at the main basin of the lake by Finnish Environment Institute.
The parameters related to the sediment transport were used as shown in Table 1. PF was applied for these four parameters.

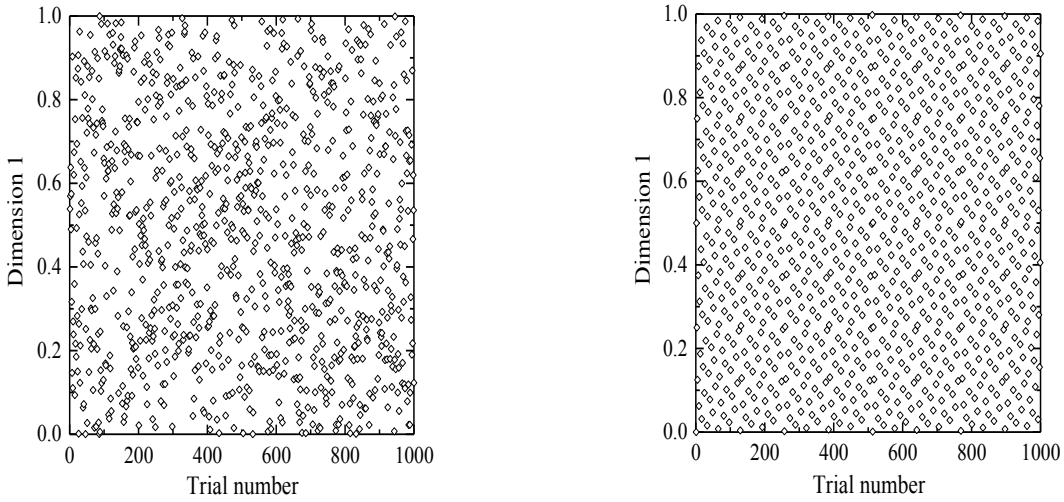
Table 1 Four parameters related to the sediment transport. Original value means one used in the model for satellite data assimilation.

Parameter name	Notes	Original value
bio_sinkrate	settling velocity [m/s]	-0.000 000 03
sedimentationrate	sedimentation rate (ex. 0.1 means that 10 % of sedimentation occurs par day.)	0.1
alphas	resuspension rate parameter [g/m2/s]	0.0025
rnsed	resuspension exponential factor	3.0

3. Methods

PF is a method for estimating probability density functions (PDFs) of state variables (Gordon, 1993; Kitagawa, 1996), and it can be applied to parameter identification of input parameters in numerical simulation. The considered sets of parameters called “particles” in PF. The simulation code of the PF was developed in (Shuku et al., 2012). Here, four parameters shown in Table 1 were regarded as “particles”, and the observation data was regarded as the reference data. Preparing a number of sets of four parameters and performing a number of model calculations from COHERENS with using the parameters (“particles”) were required for PF.

The PF usually requires about 1000 sets of parameters (particles) by a random function to find out the optimal ones. Nevertheless, by using not a random function but LDS (Low-discrepancy sequence (Tezuka, 1995)), the required sets will be reduced to 200 sets. This is because LDS generates more efficient numbers for PF to converge the results than a random function does. The images of the numbers using random function and LDS are shown in Fig. 1. Because of the time limitation, LDS was used for generating 200 particles. An application developed by one of the authors (Shuku) for generating numbers with LDS was used in this study, and the mean and the deviation for it are shown in Table 2.



(a) Random (b) LDS

Fig. 1 Image of the numbers using a random function(a) and LDS(b).

Table 2 Mean and deviation for an application with LDS for four parameters for generating 200 sets of parameters.

Parameter name	Mean	Deviation	Considered range
bio_sinkrate	-0.000 002 2	-0.000 001	-0.000 003 2 ~ -0.000 001 2
sedimentationrate	0.5	0.5	0.0 ~ 1.0
alphas	0.001 501	0.001 499	2.00E-06 ~ 0.003
rnsed	11. 5	10.505	1 ~ 22

Procedures were as follows. Some test calculations were performed with the parameters to guess the optimal parameters approximately in advance in order to narrow the range of values to be considered as particles. Here, only “bio_sinkrate” was focused because of the time limitation and that its sensitivity was the largest of the four as mentioned in the report (Mano, 2013). The tested values were in the hundredfold to hundredth range of the original value, -0.000 000 03. As the results of the test calculations, -0.000 002 20 was assumed to be close in value to the optimal one according to time series of the TSM concentration at the lake surface at the automatic station with the monitored observation data. An application for generating sets of parameters with LDS in an Excel file format was made by one of the authors (Shuku) for this study. 200 particles were generated with the setting shown in Table 1. The generated particles are shown in Fig. A-1 and listed in Table A-1. The parameters for each case were read from the file named “param.txt “(See Table A-1) automatically. This is because a directory’s name of a case was made to be equivalent to a line number in “param.txt”. Extracted time series of TSM concentration at the automatic station were gathered from 200 cases. With the data and observation data provide the optimal parameters with PF.

The calculation term was 13:00, June 26 ~ 12:00, July 6, the period when the CHL concentrations seem to be low according to satellite image and monitored the CHL

concentration at the automatic station. For the PF application, calculation term is June 26, 13:00(close to the time when satellite data was obtained, 12h59m14s) ~ 12:00, June 29. Only the results from June 28 12:00 ~ June 29 12:00 were used for running the PF application. For comparing the before-after results, the ending time was extended until 12:00, July 6. The calculation term started from June 26 for all cases, the satellite data of the day was used as the initial value of the TSM concentration on the lake surface.

4. Results

Fig. 2 shows time series of the TSM concentration extracted from 200 cases and the observation data for 24 hours. The observation data is TSM concentration with no biomass converted from Turbidity monitored at the automatic station.

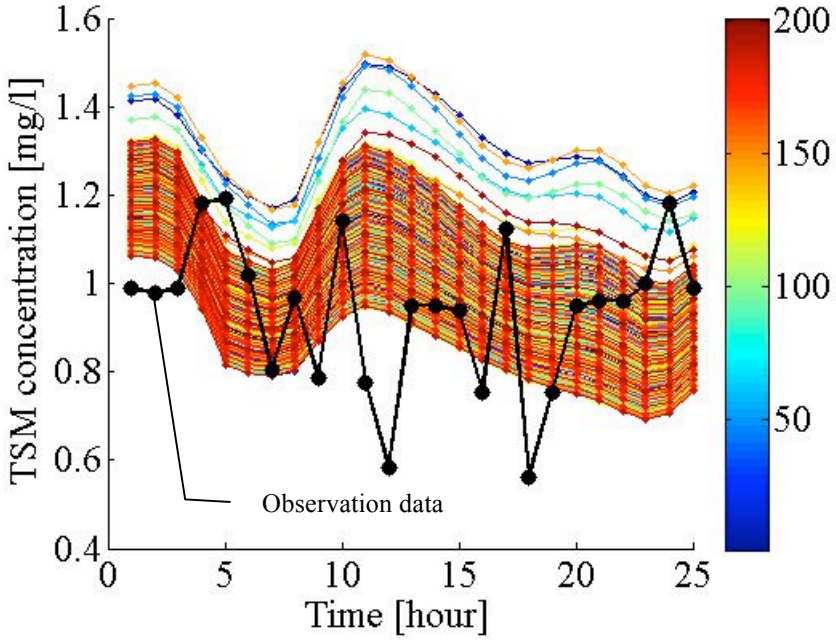


Fig. 2 Time series of TSM concentration on the lake surface at the automatic station. 200 cases of model result and an observation data.

PF was applied to the data of 200 cases and observation data. Fig. 3 shows the PF results for four parameters. The converged values show the optimal parameters. Thus the optimal values of parameters are shown in Table 3.

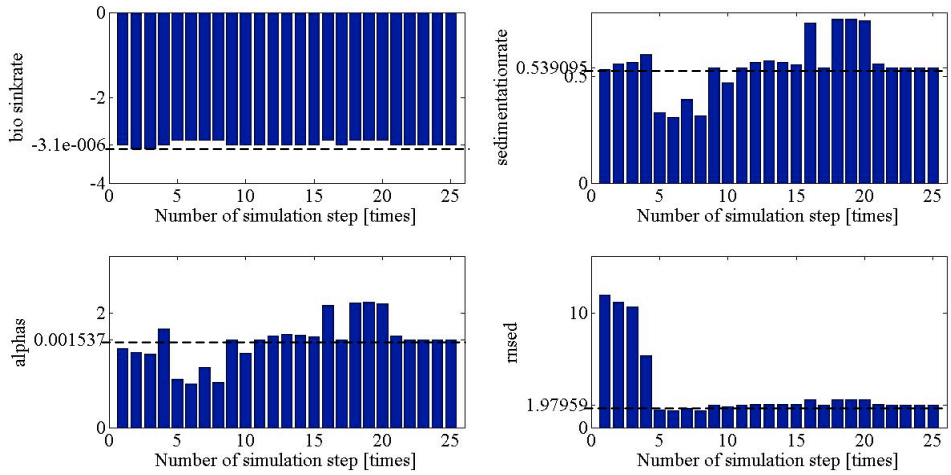


Fig. 3 PF result for four parameters

Table 3 Original values and optimal values for four parameters

Parameter name	Original value	Optimal value
bio_sinkrate	-0.000 000 03	-0.000 003 1
sedimentationrate	0.1	0.539 095
alphas	0.0025	0.001 537
rnsed	3. 0	1.979 59

5. Discussions and conclusions

Comparisons with the results before-after applying the optimal parameters were performed. Fig. 4 shows the time series of TSM concentration at the automatic station. “Before” and “After” are the results from calculations with using the original value and the optimal value in Table 3, respectively. As PF was applied to the results of “Before” from day 3 to 4 in Fig. 4 to obtain the optimal values for the four parameters, the results after day 4 shows the pure predictions. According to Fig. 4, the results of “After” are dramatically well-reproduced the observations. Thus it can be regarded that applying PF improved the model prediction in this case.

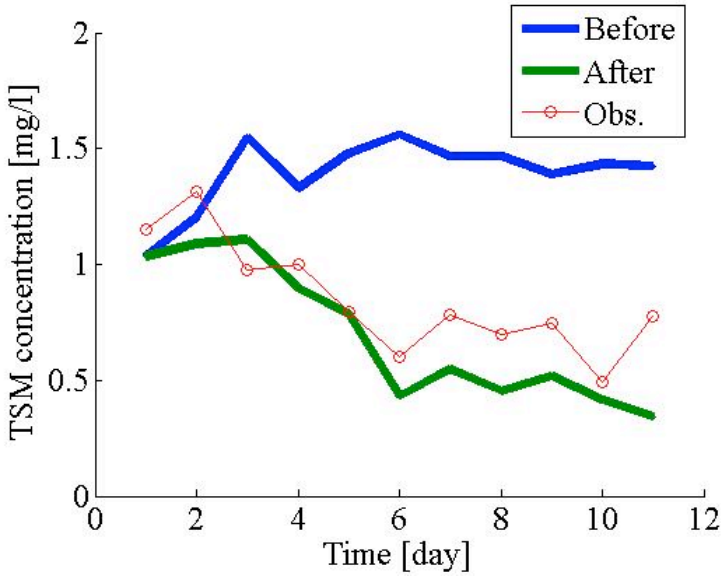


Fig. 4 Time series of TSM concentration at the automatic station.

Fig. 5 shows the comparison of vertical profile at the automatic station on 11th day of the calculation term. In the case of after applying optimal parameters, the trend, the deeper part has the larger TSM concentration, was reproduced as monitored at the automatic station (Mano, 2013).

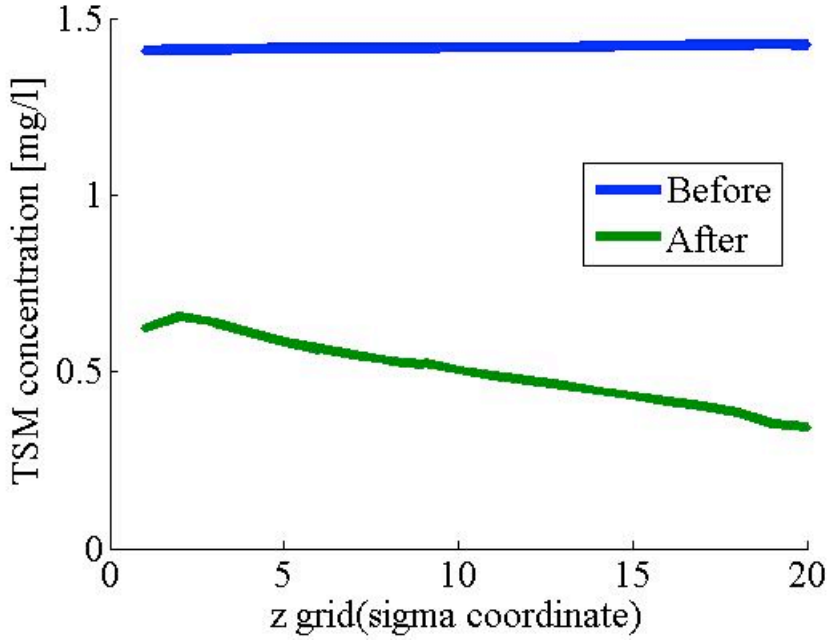


Fig. 5 Comparison of vertical profile of before-after applying the optimal parameters determined from PF on 11th day of the calculation term at the automatic station. z=20 and z=1 mean the lake surface and the bottom, respectively.

According to these results, to determine the optimal parameters with using PF seems to be very usable. This is because, not only time series at the surface but also the vertical profile, to apply PF increased the reproducibility of the model dramatically. However, to make sure the usability of PF, more tests would be required.

Two ideas for future works to make sure the usability of PF would be considered. One would be to apply PF to other calculation term, for example, not for 24 hours from June 28 12:00 but for 24 hours from on June 27 12:00. The other would be to apply PF from June 26 13:00 ~ July 6 12:00 (11 days) with using a coarser grid model. To apply PF for longer term could provide us more reliable clues for consider the usability of PF.

Acknowledgment

The authors would like to express sincere gratitude to Shin-ichi Nishimura at Okayama University and Janne Ropponen at SYKE for their guidance and kind help throughout this work.

References

[1] N. J. Gordon, D. J. Salmond, A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation", *IEE Proceedings-F*, **140**, 2, pp. 107-113, 1993.

[2] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models", *J. Compt. Graph. Statist.,* **5**, 1, pp.1-25, 1996.

[3] P. Luyten, "COHERENS — A Coupled Hydrodynamical-Ecological Model for Regional and Shelf Seas", User Documentation. Version 2.0. *RBINS-MUMM Report, Royal Belgian Institute of Natural Sciences*, **1202** pp., 2011.

[4] A. Mano, "Report for river and lake modelling group: Assimilation of remote sensing data to COHERENS V2 model of lake Sakylan Pyhajarvi", **20** pp., 2013.

[5] T. Shuku, A. Murakami, S. Nishimura, K. Fujisawa and K. Nakamura, "Parameter identification for Cam-Clay model in partial loading model tests using the particle filter", *Soils and Foundations*, **52** 2, pp. 279-298, 2012.

[6] S. Tezuka, "Uniform random numbers: Theory and Practice", *Kluwer Academic Publishers*: Boston, 1995.

Assimilation of satellite data to 3D hydrodynamic model of Lake Säkylän Pyhäjärvi

Akiko Mano[1], Olli Malve[2], Sampsa Koponen[2], Kari Kallio[2], Antti Taskinen[2], Janne Ropponen[3], Janne Juntunen[3], Ninni Liukko[3]

[1] Ochanomizu University, 2-1-1, Ohtsuka, Bunkyo-ku, Tokyo 112-8610, Japan, mano.akiko@ocha.ac.jp

[2] Finnish Environment Institute, P.O. Box 140, SF00251 Helsinki, Finland

[3] Finnish Environment Institute, Technopolis, Survontie 9, 40500 Jyväskylä, Finland

Abstract:

To analyze the applicability of occasional initialization (OI) of total suspended matter (TSM) concentration field based on turbidity derived from satellite data to numerical simulation, dispersion studies of suspended matter in Lake Säkylän Pyhäjärvi (lake area 154 km^2; mean depth 5.4 m) were conducted using the 3D COHERENS simulation model. To evaluate the practicality of OI, five cases with different initialization frequencies were conducted: 1) every time, when satellite data were available; 2) every 10 d; 3) 20 d; and 4) 30 days; 5) control run without repeated initialization. To determine the effectiveness of initialization frequency, three methods of comparison were used: simple spatial differences of TSM concentration without biomass in the lake surface layer; averaged spatial differences between initialization data and model results; and time series of TSM concentration and observation data at 1 m depth at the deepest point of the lake. Results showed that OI improves the prediction significantly when applied often, i.e. at 10 day intervals. Even if it is applied less often, OI improves the prediction.

Key words

COHERENS; model prediction; suspended particulate matter; suspended solids; turbidity; water quality

This paper has already been submitted to water science and technology

WP 4

Statistical analysis for habitat modelling and relationship between land use and river water quality

Juvenile salmon patch identification and comparison using Echelon analysis

Makiko Oda[*], Saija Koljonen[**], Fumio Ishioka[†], Petteri Alho[††], Hiroshi Suito[†], Timo Huttula[**] and Koji Kurihara[†]

[*] National Defense Medical College
3-2 Namiki, Tokorozawa, Saitama, 359-8513, Japan
oda@ndmc.ac.jp
[**]Finnish Environment Institute
Saija.Koljonen@ymparisto.fi, timo.huttula@ymparisto.fi
[†] Okayama University, Japan
fishioka@law.okayama-u.ac.jp, suito@okayama-u.ac.jp, kurihara@ems.okayama-u.ac
[††] Department of Geography and Geology, University of Turku
mipeal@utu.fi

Abstract:

We studied a habitat patch identification technique for juvenile Atlantic salmon utilizing the suitability index and Echelon analysis. After identification of patches, we evaluated each identified patch using the likelihood ratio statistic so that the best suitability areas for salmon were determined. Laser and sonar are interferometric measuring systems we used for our salmon data. As a result of comparison between these two data sets, we found that sonar was more optimistic than laser for measuring suitability index.

1. Introduction

Various patch identification methods for habitats have been published (Fortin et al., 1995; Plotkin et al., 2002). A patch is defined as a spatially homogeneous area where at least one variable has similar attributes either of category or quantitative value (Fortin et al., 2005). Therefore, a patch was adopted in studies for both plants and animals. Patches can be identified using various variables such as tree or animal abundance and percentage coverage of trees. However, a solid method of identifying patches has not been established yet.

In the present study, we identified juvenile salmon patches using Echelon analysis. Oda et al. (2012) suggested the technique for identification of patches in a forest using Echelon analysis. We presented that the technique can also be used for an animal, juvenile salmon, and evaluated the identification patches.

In this paper, we first explained the patch identification method we used, and identified juvenile salmon patches based on the suitability index. We then assessed these identified

patches using the two different patch evaluation methods called laser and sonar, and finally compared the data derived from these two patch evaluation methods for measuring suitability index.

2. Survey site and data

The subarctic river Utsjoki is a tributary of a highly productive Atlantic salmon (*Salmo salar*, L.) river, Tana River. We modeled a stream stretch, which is known to have a strong Atlantic salmon juvenile population. The modeling procedure was undertaken for two different datasets with different accuracy: laser and sonar.

We applied life stage-specific habitat preference criteria (HPC) for depth, velocity and substrate preference by Atlantic salmon juvenile (Mäki-Petäys et al., 2002, 2004, unpublished data at http://www.rktl.fi/www/uploads/pdf/raportti284.pdf) to acquire a combined suitability index (CSI) value (product of habitat preference values). For this analysis we used only one discharge situation (20 m^3s^{-1}) recognized as a normal summer flow.

3. Echelon analysis

Echelon analysis (Myers et al., 1997; Kurihara et al., 2000; Ishioka et al., 2007) is a method to show phase structure in Echelon dendrogram based on response variables and neighboring information.

3.1. Echelon analysis procedure

A phase structure of data can be shown using Echelon dendrogram (Figure 3.1): The surface value is shown such as contour lines in the left picture in Figure 3.1. The middle and right pictures show a lateral view. The right picture is Echelon dendrogram.

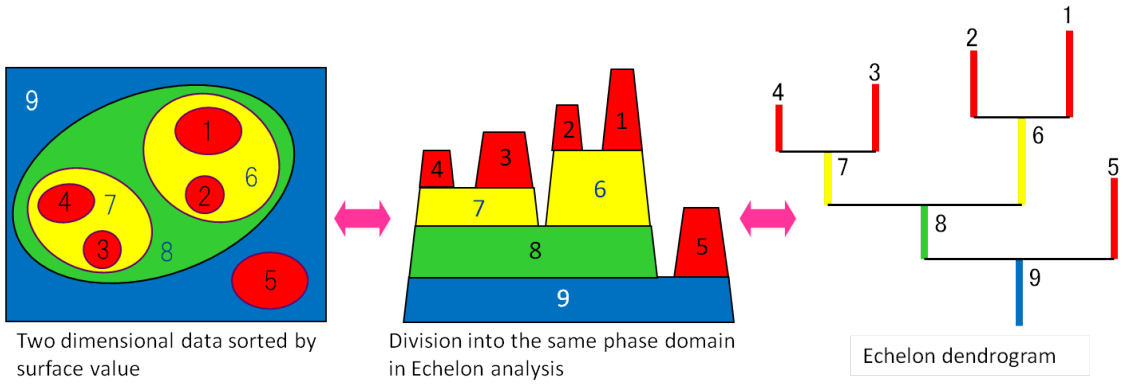


Fig. 3.1 The diagram of Echelon analysis.

We call the Echelon numbers 1, 2, 3, 4 and 5 as a “peak”, the Echelon numbers 6, 7 and 8 as a “foundation” and the Echelon number 9 as a “root”.

Each cell has a value as presented in Figure 3.2, and each cell has neighboring information. For example, cell [C3] is located at the center of its adjacent cells of [B2], [C2], [D2], [B3], [D3], [B4], [C4] and [D4].

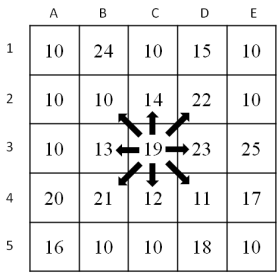


Fig. 3.2 5×5 array data.

1. Establish peaks

   At first, the cell [E3], with the maximum value in the 5×5 array data, is identified (Figure 3.3). Among adjacent cells of the cell [E3], the largest value, the cell [D3] is identified. The third ([D2]) and fourth ([C3]) values are identified in the same manner. The cell [C3] has the adjacent cell [B4], which is larger than the cell [C3]. Therefore, the first peak consists of three cells ([D2], [D3], [E3]). Other peaks {G(2), G(3), G(4)} are found in same manner (Figure 3.4).

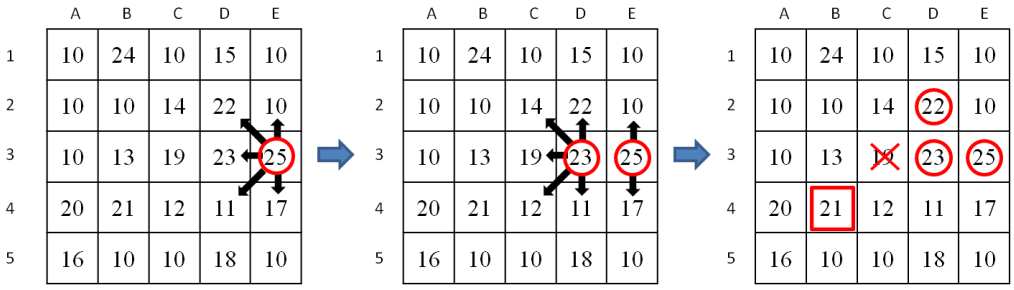


Fig. 3.3 Established peaks procedure.

2. Establish foundations

   Excluding the four peaks identified earlier, the cell [C3] is the maximum value in 5×5 array data based on its spatial location relationship (Figure 3.5), followed by the cell [E4]. Next, the cell [C2] becomes foundation between the G(2) and foundation [E4], [A5] and [D1]. The rest of cells are included in same [C2] foundation, since these are inevitably adjacent to peaks and the foundations.

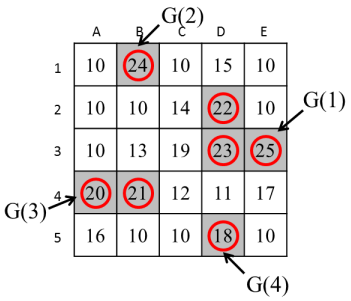


Fig. 3.4 Established peaks.

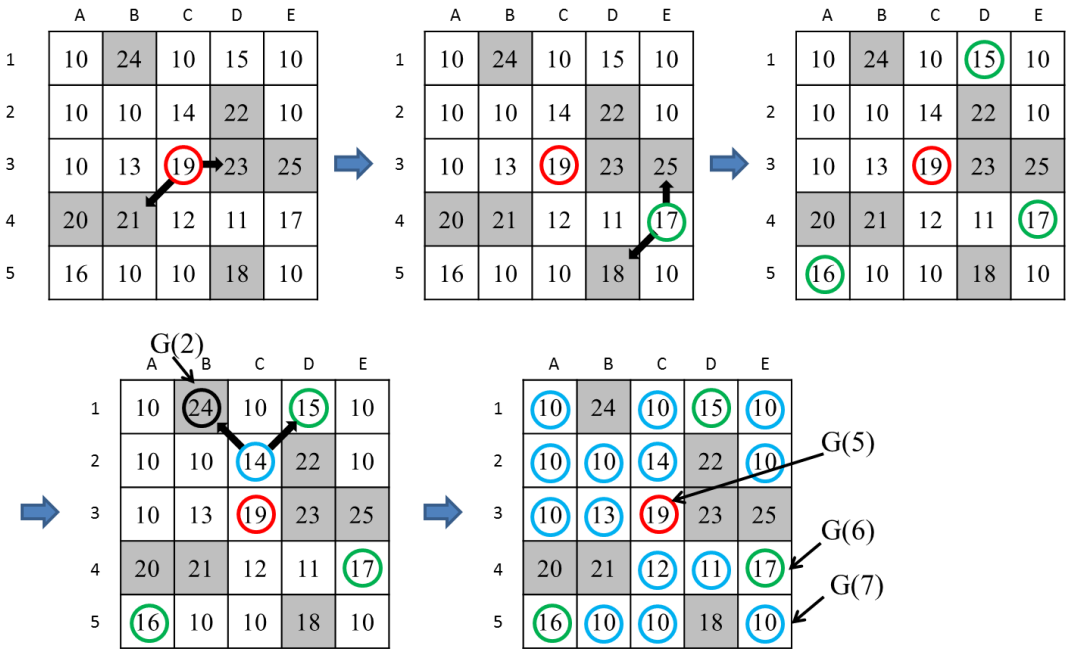


Fig. 3.5 Established foundations.

3. Establish Echelon dendrogram

Echelon dendrogrm can be established when peaks and foundations are identified in 5×5 array data. Four identified peaks: G(1) – G(4) are linked to foundations in order to draw a dendrogram (Figure 3.6). The foundation G(5) links the peaks G(1) and G(3), and the cell E[4] links the peak G(4) and foundation G(5). The cells [A5] and [D1] belong to the foundation of cell [E4]. These three cells establishes the foundation G(6). The rest of cells are the root G(7).

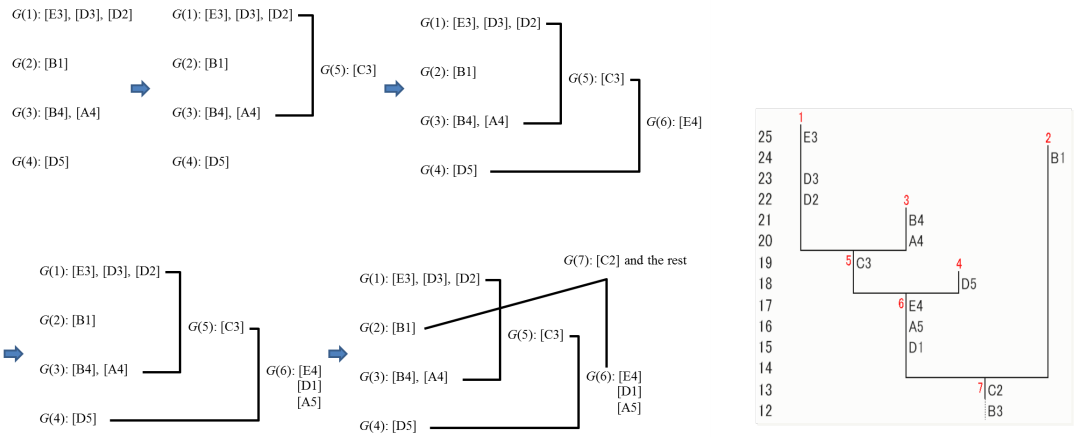


Fig. 3.6 Echelon dendrogram procedure.

2.2. Hotspot Detection

Establishing an Echelon dendrogram makes it easy to detect statistical significance among groups like G(1). This is an advantage of Echelon analysis. We call this significant area "hotspot". There is a subset area Z in the area G. p_1 is the population probability with an attribute within the area Z, and p_2 is the population probability with an attribute outside the area Z. The probabilities of all individuals with attributes are mutually independent. The hypothesis is as follows:

$$H_0 : p_1 = p_2 = p \qquad v.s. \qquad H_1 : p_1 > p_2$$

Where $n(G)$ is the total population in the area G, $n(Z)$ is the population within the area Z, $c(G)$ is the number of attributes in the area G, and $c(Z)$ is the number of the attributes within the area Z (Figure 3.7). Poisson model is used.

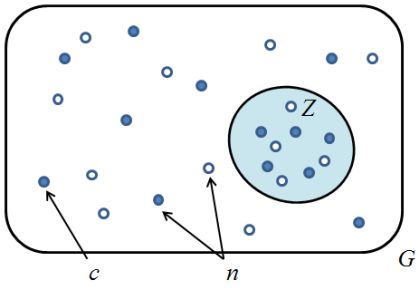


Fig. 3.7 Population n and the number of the attribute c in area G. Comparing $p_1 = c(Z)/n(Z)$ and $p_2 = c(G)/n(G)$.

The probability of the number of points is $c(G)$ in the area G is

$$\frac{\exp[-p_1 n(Z) - p_2 n(\bar{Z})][p_1 n(Z) + p_2 n(\bar{Z})]^{c(G)}}{c(G)!} \qquad (3.1)$$

The density at location x in the area G is

$$\frac{p_1 n(x)}{p_1 n(Z) + p_2 n(\bar{Z})} \quad \text{if} \quad x \in Z \qquad (3.2)$$

$$\frac{p_2 n(x)}{p_1 n(Z) + p_2 n(\bar{Z})} \quad \text{if} \quad x \in \bar{Z} \qquad (3.3)$$

The likelihood function of Poisson model is

$$\begin{aligned} L(Z, p_1, p_2) &= \frac{\exp[-p_1 n(Z) - p_2 n(\bar{Z})][p_1 n(Z) + p_2 n(\bar{Z})]^{c(G)}}{c(G)!} \\ &\times \prod_{x_i \in Z}^{n_i \in Z} \frac{p_1 n(x)}{p_1 n(Z) + p_2 n(\bar{Z})} \prod_{x_i \in Z^c}^{n_i \in Z^c} \frac{p_2 n(x)}{p_1 n(Z) + p_2 n(\bar{Z})} \qquad (3.4) \\ &= \frac{\exp[-p_1 n(Z) - p_2 n(\bar{Z})}{c(G)!} p_1^{c(Z)} p_2^{c(\bar{Z})} \prod_{x_i}^{n} n(x_i) \end{aligned}$$

Let $x(Z)$ be a random variable for the number of attributes within the area Z. Anywhere under the area Z,

$$x(A) \sim Poisson(p_1 n(A \cap Z) + p_2 n(A \cap \bar{Z})) \qquad (3.5)$$

Under the null hypothesis,

$$x(A) \sim Poisson(pn(A)) \tag{3.6}$$

To maximize the likelihood function, the maximum likelihood function over the area Z is calculated. The maximum likelihood estimators $\hat{p}_1 = c(Z)/n(Z)$ and $\hat{p}_2 = c(\overline{Z})/n(\overline{Z})$ are estimated.

$$L(Z) = \frac{\exp[-c(G)]}{c(G)!}\left(\frac{c(Z)}{n(Z)}\right)^{c(Z)}\left(\frac{c(\overline{Z})}{n(\overline{Z})}\right)^{c(\overline{Z})}\prod_{x_i}^{n} n(x_i) \tag{3.7}$$

The likelihood ratio $\lambda(Z)$ is the maximum value in the subset area within the area G to detect hotspots.

$$\lambda(Z) = \frac{\max_Z L(Z)}{L_0} = \frac{\left(\frac{c(Z)}{n(Z)}\right)^{c(Z)}\left(\frac{c(\overline{Z})}{n(\overline{Z})}\right)^{c(\overline{Z})}}{\left(\frac{c(G)}{n(G)}\right)^{c(G)}} \tag{3.8}$$

L_0 is the following likelihood function under the null hypothesis,

$$L_0 := \sup_p \frac{\exp[-pn(G)]}{c(G)!} p^{c(G)} \prod_{x_i}^{n} n(x_i) = \frac{\exp[-c(G)]}{c(G)!}\left(\frac{c(G)}{n(G)}\right)^{c(G)}\prod_{x_i}^{n} n(x_i) \tag{3.9}$$

The test statistic $\lambda(Z)$ can be

$$\lambda(Z) = \left(\frac{c(Z)}{e(Z)}\right)^{c(Z)}\left(\frac{c(\overline{Z})}{e(\overline{Z})}\right)^{c(\overline{Z})} \tag{3.10}$$

where $e(Z)$ is the expectation of the attribute within the area Z, and $e(G)$ is equal to $c(G)$.

2.3 Hotspot detection procedure using Echelon dendrogram

The hotspot detection of Echelon analysis is as follow:

1. Draw the Echelon dendrogram for target data.
2. Scan the areas from upper Echelon to the bottom, based on the hierarchical structure determined in Step 1.
3. Detect the hotspot, which takes the maximum log likelihood ratio; $\log\lambda(Z)$.

The significance of the hotspot candidate is evaluated using Monte Carlo simulation.
A p-value based on Monte Carlo simulation can be obtained as follow:

1. Generate a random data set under the null hypothesis when we condition with conditions on the total number of attribute $c(G)$.
2. Calculate the $\log\lambda(Z)$ from the simulated data.
3. Repeat a process of Step 1 and 2 multiple times.

4. Define as $p = R/(\sharp SIM + 1)$

where R is the rank of the test statistics from the real data set among all data sets and $\sharp SIM$ is the number of simulated data sets that have been generated.

Echelon analysis detects two hotspot candidates: most likely cluster and secondary cluster. The most likely cluster is the highest $\log\lambda(Z)$, and the secondary cluster $\log\lambda(Z)$ is the second highest.

4. Patch identification method

4.1. Patch identification method

First, an area with significantly high suitability was detected based on the Echelon analysis hotspot detection method (Figure 4.1 and 4.2, Table 4.1) and salmon suitability index. The vertical line of Echelon dendrogram is the suitability index. If the number of areas is too large, we can set the maximum number of hotspot areas in advance. This maximum number is K. K is the maximum number of areas in most likely cluster. The secondary cluster can also be detected in the same manner. Here, K=500 (approximately 10% of the total) was adapted (Figure 4.3). Detected areas included not good suitability (<0.5). Therefore, we proposed another patch identifying method (Figure 4.4). The top ten patches of maximum suitability within each patch area were shown (Figure 4.5 and 4.6).

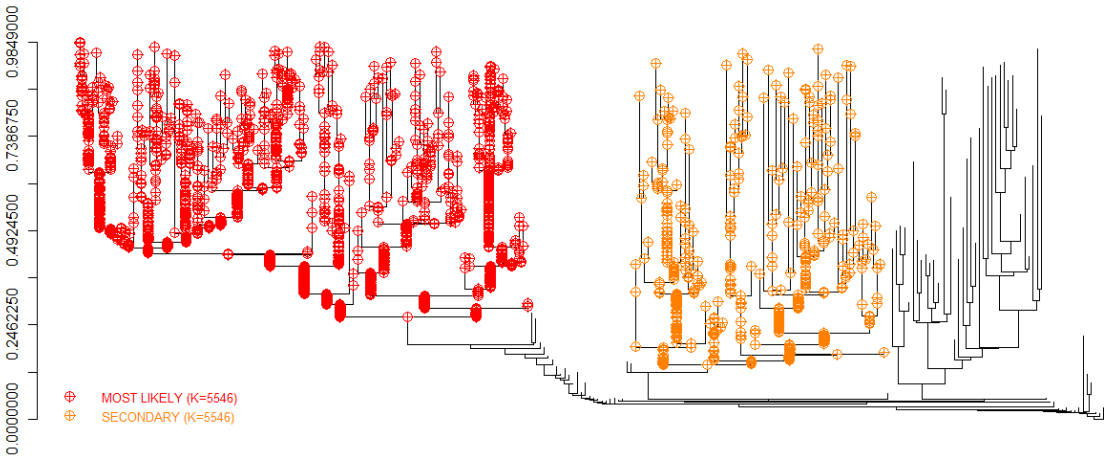


Fig. 4.1 Echelon dendrogram (laser) marked significantly high areas.

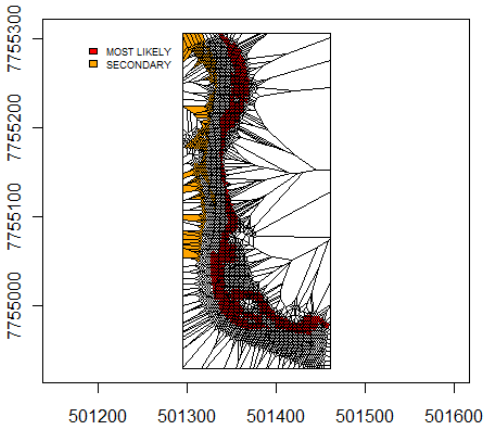


Fig. 4.2 Significantly high areas (laser).

Table 4.1 Significantly high areas (laser).

	Most likely	**Secondary**
The number of area	1707	489
logλ	383.07	19.49
p-value	0.001	0.001

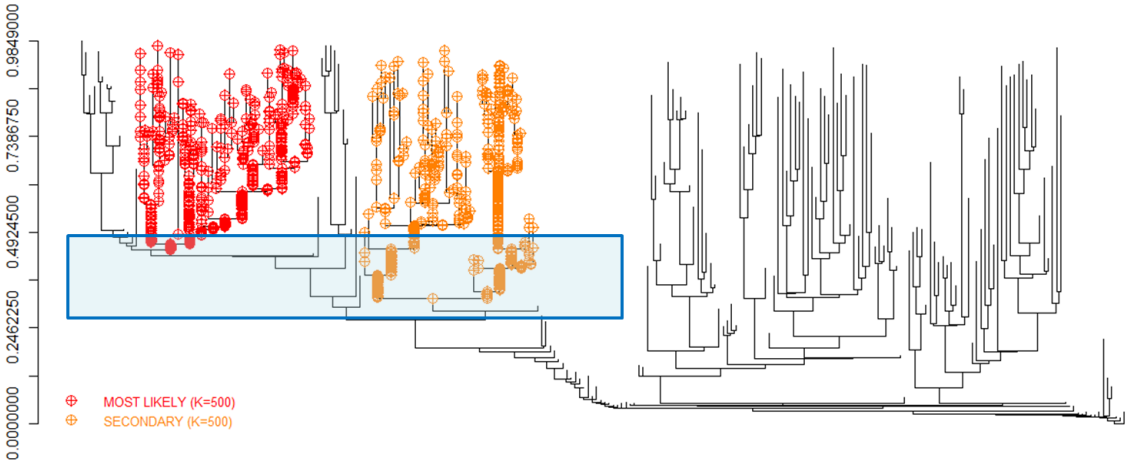


Fig. 4.3 Echelon dendrogram (laser) marked significantly high areas (*K*=500).

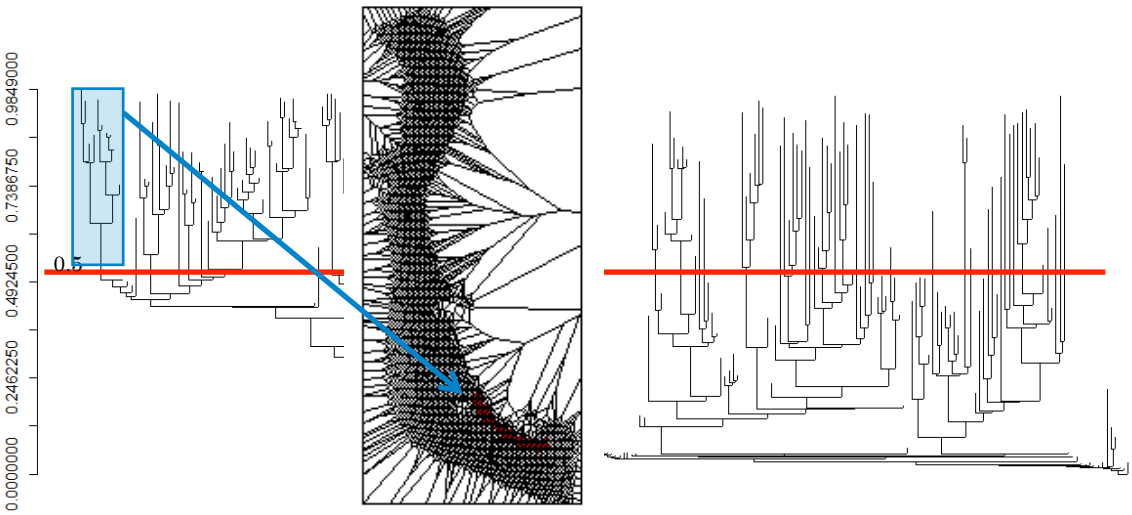


Fig. 4.4 Proposed patch identifying method.

4.2. Patch evaluation

Some identified patches are too small, therefore a patch size criterion or patch evaluation were needed. Eq. (3.10) can be useful as patch evaluation. The result is shown in Table 4.2 and 4.3. Obviously, logλ of the too small patch was small, therefore these logλ were utilized to evaluate patches.

Table 4.2 The number of cell and logλ (laser).

ID	1	2	3	4	5	6	7	8	9	10
The number of cell	10	102	770	4	79	2	12	81	22	10
logλ	42.61	22.92	208.08	0.78	17.51	0.51	2.91	13.22	5.58	1.92

Table 4.3 The number of cell and logλ (sonar).

ID	1	2	3	4	5	6	7	8	9	10
The number of cell	158	80	61	2	9	267	140	24	2	19
logλ	44.79	23.18	15.56	0.64	2.84	75.44	33.18	6.81	1.02	6.13

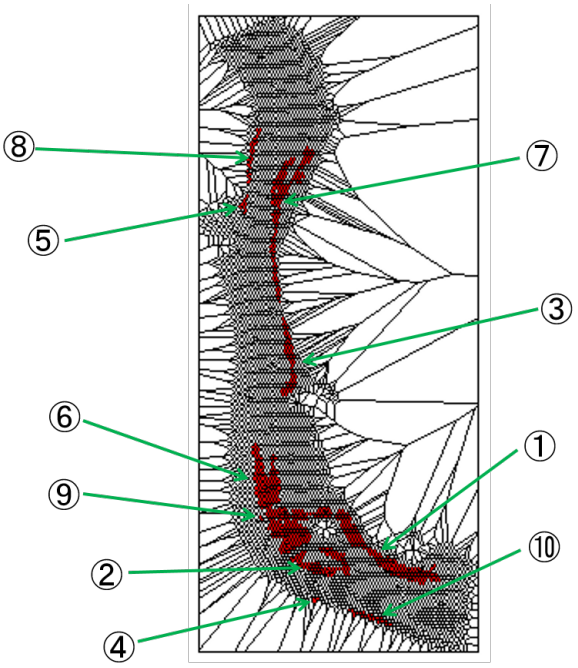


Fig. 4.5 Identified ten patches (laser).

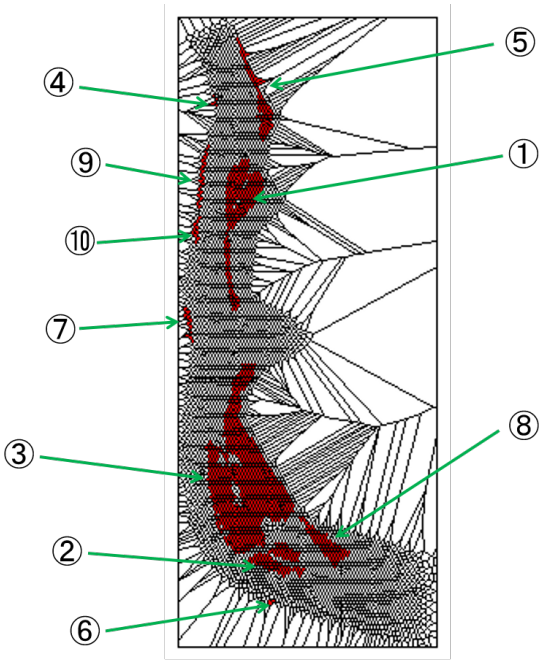


Fig. 4.6 Identified ten patches (sonar).

5. Comparing two kind of data: laser and sonar

logλ are indexes evaluated only in same area. Laser data and sonar cannot be compared using logλ. An advantage of Echelon analysis is easy scan. We can compare two kinds of data by making a dendrogram from both laser and sonar. A conventional method is detecting significance area in each area, it cannot compare some areas. New method can assess same area as one area and one dendrogram (Figure 5.1). The hypothesis is as follows:

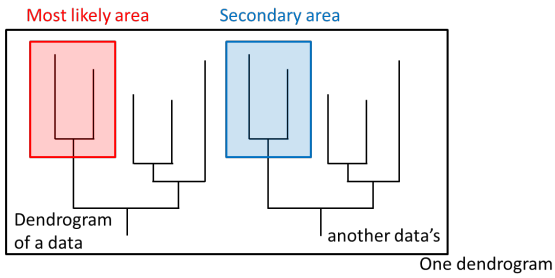


Fig. 5.1 The Echelon dendrogram with two kind of data method.

H_0 : All area's suitability indexes are equal.
H_1 : not H_0

The result is Figure 5.2. The first and forth hotspot were sonar data, second and forth were laser data. This result might have been related that the laser measurement accuracy was higher than the sonar.

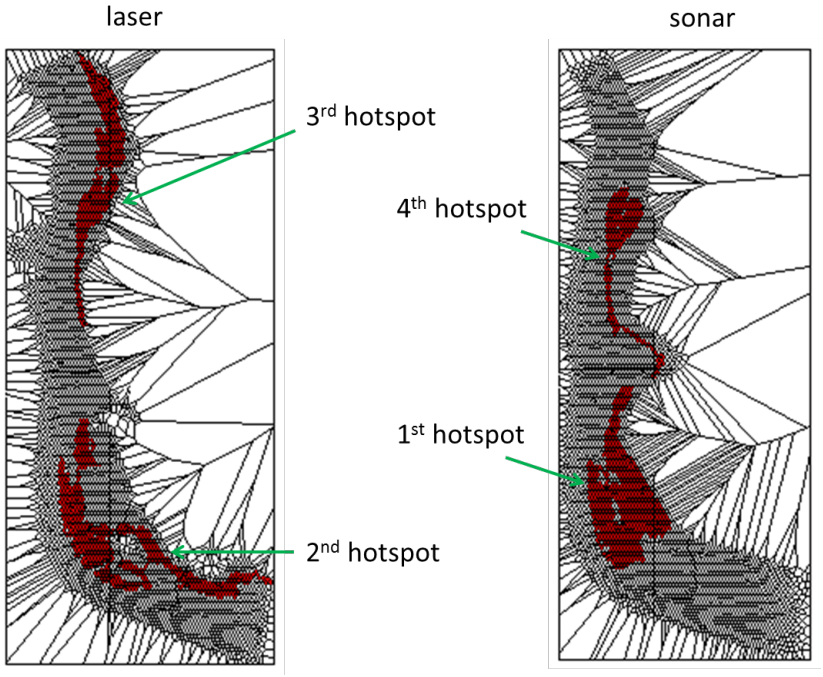


Fig. 5.2 Detected hotpot areas (p-value<0.5).

6. Discussion

In this study, we confirmed that the method of patch identification was effective in salmon habitat suitability. The identification of patches through Echelon dendrogram and salmon suitability index was more useful than detecting significant areas from spatial scan statistic. It is possible to show the best habitat for salmon by evaluating patches. However, in previous finding (Vadas and Orth, 2001), suitability index ≥ 0.75 were defined as optimal. We should further examine the critical suitability index. We also presented comparison method with two data (sonar and laser) using Echelon analysis. We want to establish the comparing method because it has much room to study.

References

[1] A. Mäki-Petäys, A. Huusko, J. Erkinaro and T. Muotka, "Transferability of habitat preference criteria of juvenile Atlantic salmon (*Salmo salar*) ", *Canadian Journal of Fisheries and Aquatic Science*s, **59**, pp. 218–228, 2002.

[2] A. Mäki-Petäys, J. Erkinaro, E. Niemelä, A. Huusko and T. Muotka, "Spatial distribution of juvenile Atlantic salmon (*Salmo salar*) in a subarctic river: size-specific changes in a strongly seasonal environment", *Canadian Journal of Fisheries and Aquatic Sciences*, **61**, 12, pp.2329–2338, 2004.

[3] F. Ishioka, K, Kurihara, H. Suito, Y. Horikawa and Y. ONO, "Detection of hotspots for three-dimensional spatial data and its application to environmental pollution data", *Journal of Environmental Science for Sustainable Society*, **1**, pp.15–24, 2007.

[4] J. B. Plotkin, J. Chave and P. S. Ashton, "Cluster analysis of spatial patterns in Malaysian tree species", *The American Naturalist*, **160**, pp. 629–644, 2002.

[5] K. Kurihara, W. L. Myers and G. P. Patil, "Echelon analysis of the relationship between population and land cover patterns based on remote sensing data", *Community Ecology*, **1**, pp. 103-122, 2000.

[6] M.-J. Fortin and M. R. T. Dale, *Spatial Analysis: A guide for Ecologists*. New York: Cambridge, University Press, 2005.

[7] M.-J. Fortin and P. Drapeau, "Delineation of ecological boundaries: comparison of approaches and significance tests", *Oikos*, **72**, pp. 323-332, 1995.

[8] M. Oda, F. Ishioka, T. Masaki and K. Kurihara, "Forest partition using Echelon hierarchical structure", Bulletin of Data Analysis of Japanese Classification Society, **1**, 2, pp.17-31, 2012.

[9] R. L. Vadas, Jr and D. J. Orth, "Formulation of habitat suitability models for stream fish guilds: do the standard methods work?", *Transactions of the American Fisheries Society*, **130**, pp. 217-235, 2001.

[10] W. Myers, G. P. Patil and K. Joly, "Echelon approach to areas of concern in synoptic regional monitoring", *Environmental and Ecological Statistics*, **4**, pp. 131–152, 1997.

Water quality affects diatom diversity in Finnish rivers

Makiko Oda[*], Matti Leppänen[†], Anna Karjalainen[†], Satu-Maaria Karjalainen[†], Hiroshi Suito[**] and Timo Huttula[†]

[*] National Defense Medical College
3-2 Namiki, Tokorozawa, Saitama, 359-8513, Japan
oda@ndmc.ac.jp
[†]Finnish Environment Institute
Matti.T.Leppanen@ymparisto.fi, Anna.Karjalainen@ymparisto.fi,
Satu.Maaria.Karjalainen@ymparisto.fi, timo.huttula@ymparisto.fi
[**]Okayama University, Japan
suito@okayama-u.ac.jp

Abstract:
Land use has a great impact on water quality, which can affect diatom diversity and ecology. We assessed statistically these relationships using a spatial method. As the results, using classification and regression trees, we demonstrated low pH having a great impact on diatom diversity.

1. Introduction

Technical development has made a study based on spatial information much easier. This type of study technique is utilized in various disciplines such as Population Biology (Bellier et al., 2013) and Epidemiology (Takahashi et al., 2008). A study about water quality also uses spatial information. LaBeau et al. (2014) presented relationship between land use and phosphorus (P) and estimated the amount of P loading.

We studied the impact of water quality on diatoms using land use data. We found the spatial relationship between water quality and diatom utilizing Echelon analysis. Then, we assessed the impact of water quality on diatoms using classification and regression trees (CART).

In this paper, we explained the survey area, data and analysis methods followed by the results of Echelon analysis and CART. We also identified the substances that impact on diatoms.

2. Materials and methods

Figure 2.2 shows our study procedure. We analyzed land use data, water quality data and diatoms data utilizing Echelon analysis and CART.

2.1. Study area

The study area was the mid-western part of Finland (Figure 2.1). The water quality data were collected from the 13 catchment areas (rivers) with land use data and had altogether 29 sampling sites.

2.2. Land use data

Land use data was measured in 2006. Land use of each catchment area was classified in 43 types, and they were grouped into ten classes based on soil characteristics (Table 2.1) because of too detailed to estimate a missing value.

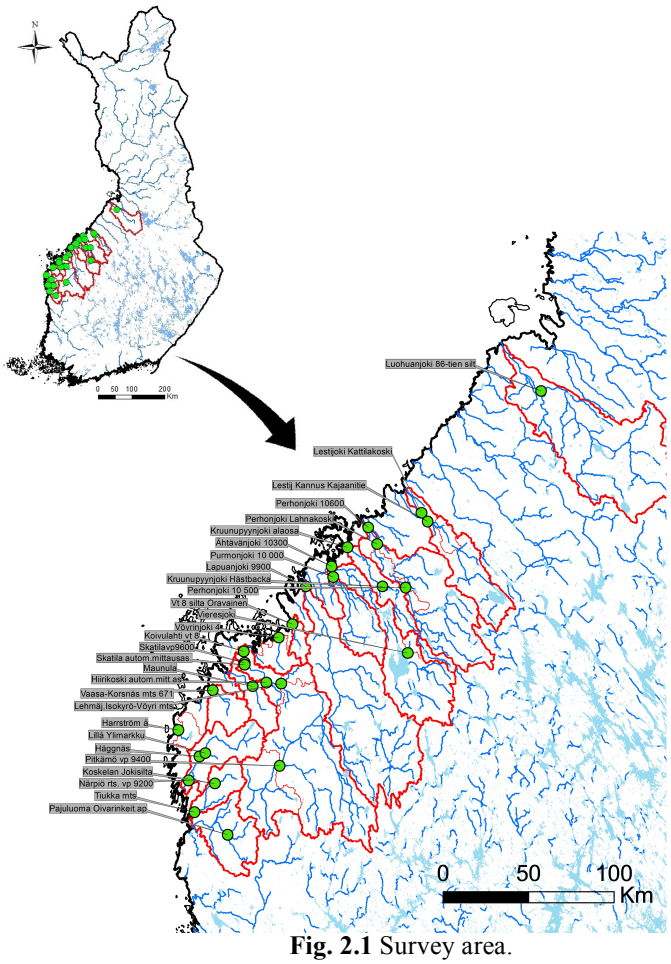


Fig. 2.1 Survey area.

Table 2.1 Ten classes of land use grouped based on soil.

Class of land use	Original land use
Artificial area	Density constructed urban areas, sparsely constructed urban areas, industrial and service areas, traffic areas, harbor areas, airport areas, area for land taking, disposal sites, under construction, parks and green areas in urban areas, summer cottages, sport areas
Agricultural area	Fields, fruit tree and berry production areas, stock feeding areas, small agricultural mosaic
Mineral soil area	Deciduous forests on mineral soil, evergreen forests on mineral soil, mixed forest on mineral soil, rare tree areas, cc <10-30% on mineral ground
Peat area	Deciduous forests of peat land, evergreen forests on peat land, mixed forest on peat land, rare tree areas, cc <10-30% on peat ground, peat production areas
Rocks area	Evergreen forests on rocks, mixed forest on rocks, rare tree areas, cc <10-30% on rock ground, rock ground

Wet area	Moor land with rare growing, inland wetlands on the land, inland wetlands in the water, open swamps, wetlands on the land area close to the sea, wetlands in the water close to the sea
Others	Natural grasslands, thin tree groups and moors, rare tree areas, cc <10%, rare tree areas over boreal forest, rare tree areas, cc <10-30% on below electricity line, sandy beaches
River	River
Lake	Lake
Sea	Sea

2.3. Water quality data

Water quality was surveyed from 2006 to 2012, and the number of sampling sites was 29. However, none of the sites had complete data (i.e. without missing variables). The Table 2.2 shows the percentages of missing variables pertaining to the water quality sampling sites.

Table 2.2 The percentages of missing variables.

Substance	Alkalinity	Al	Cd	COD	Particulate matter	Total S	Cr
Missing value rate (%)	51.3	76.9	79.9	59.0	70.7	100	80.2

Cu	Pb	Ni	Total Organic Carbon	Fe	Turbidity	Zn	SO4	Color	pH
80.2	80.2	79.9	92.3	94.1	63.0	80.2	95.6	51.3	49.1

2.4. Diatom data

Diatoms state was evaluated using two methods: number of type-specific species (TT_{40}) and percent model affinity (PMA). Both TT_{40} and PMA are between 0 and 1: close to 0 means the bad state, close to 1 means the best.

TT_{40} (Aroviita et al., 2008) is a taxon type specific index, which tells about how many taxa are found in the studied site compared to the reference site. The higher the index value, the higher the number of the same taxa in the sites. A low index value shows that there is a difference in the diatom community at the study site compared to a reference community.

PMA (Novak and Bode, 1992) describes the similarity of the reference and studied diatom community. If the index value is high, the communities in the studied and reference environment area are the same. If the value is low, the community in the studied site differs from the one in the reference site.

These data were surveyed in summer from 2007 to 2012, and the number of survey areas

varied from year to year. 2010 was the highest number of survey with 17 areas, and the lowest was 2008 with no survey taken place.

2.5. Models for land use data and water quality

Water quality data had multiple missing values; therefore, we made the complete data using two kinds of method; near sampling site data and estimating generalized linear model (GLM). We preferably used the sampling site data from the same catchment area, near sites and the same river. Nevertheless, if data had a missing value, we estimated them using GLM according to equation 2.1.

$$Y_j = \beta_0 + \sum_i \beta_i X_i + \varepsilon \qquad (2.1)$$

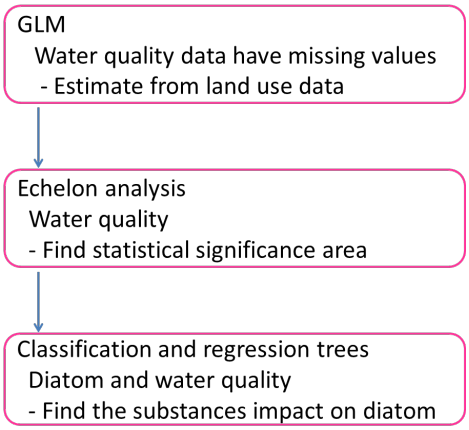


Fig. 2.2 Schematic diagram.

where Y_i is a variable like alkalinity and X_i are year, season and each land area like artificial area or agriculture area of Table 2.1.

However, the estimation the substance like metals was not successful due to too many missing values. Therefore, we estimated the missing values of alkalinity, COD, turbidity and pH.

2.6. Spatial model for water quality

Establishing an Echelon dendrogram makes it easy to detect statistically significant areas. For additional information, see a section 3 of "Juvenile salmon patch identification and comparison using Echelon analysis".

We detected hotspot areas from 3D data based on time series data, therefore time (season and year) and areas information were included in significantly high areas. Statements of significance are based on $p \leq 0.05$.

2.7. Models for water quality and diatoms

Classification and regression trees (CART) were obtained by recursively partitioning the data space and fitting a simple prediction model within each partition. As a result, the partitioning can be represented graphically as a decision tree (Loh, 2011).

A variable was divided based on impact on the response variable. In this case, response variables were the diatom indexes and explanatory variables were the water quality data.

3. Results

3.1 Hotspot areas of alkalinity, COD, pH and turbidity

We detected hotspot areas of alkalinity, COD, pH and turbidity using an Echelon analysis. Each Echelon dendrogram is shown in Figures 3.1, 3.2, 3.3 and 3.4, and Tables 3.1, 3.2, 3.3 and 3.4 explain the hotspot areas. The red symbols represent the most likely significance areas and the orange symbols represent the secondary likely areas in each figure. The *p*-values of all the most likely hotspot areas and secondary areas were less than 0.5; therefore, the detected areas were the statistically significant areas. *K* in figures is the maximum number of areas in the most likely cluster. The secondary cluster can also be detected in the same manner. Here, *K*=27 (approximately 10% of the total) was adapted. However, the number of diatom samples was 47.

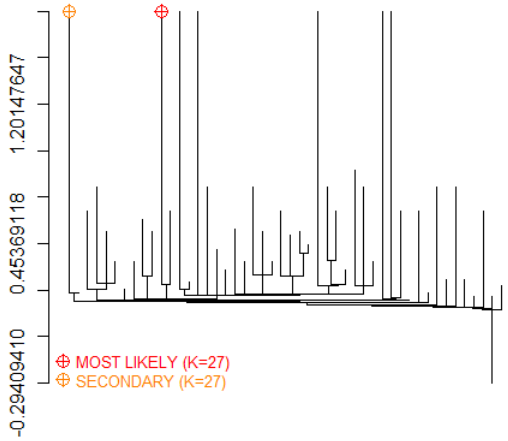


Fig. 3.1 The Echelon dendrogram of alkalinity (2006-2012).

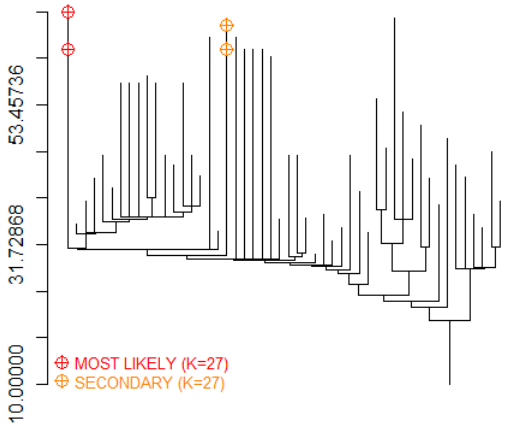


Fig. 3.2 The Echelon dendrogram of COD (2006-2012).

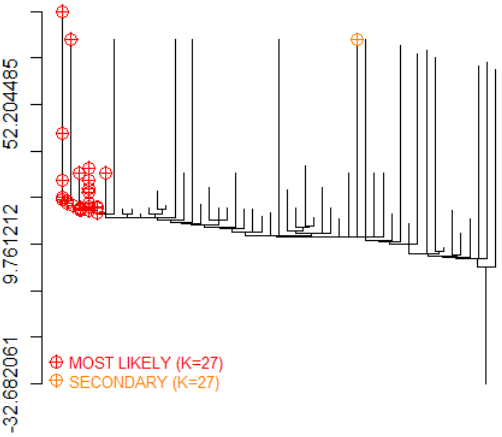


Fig. 3.3 The Echelon dendrogram of turbidity (2006-2012).

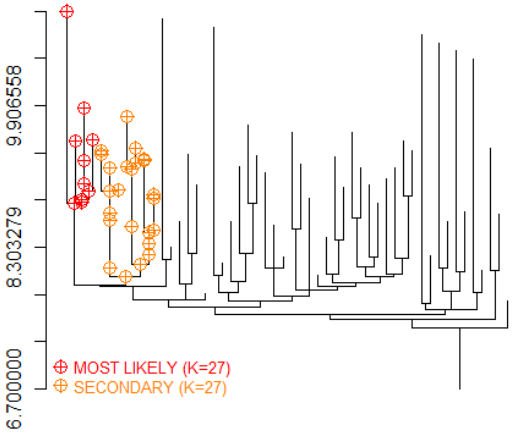


Fig. 3.4 The Echelon dendrogram of low pH (2006-2012).

Table 3.1 The Hotspot areas of alkalinity (2006-2012).

	p-value	season	year	sampling site
hotspot	0.001	summer	2006	Purmoj.10k
secondary	0.001	summer	2012	Purmoj.10k

Table 3.2 The Hotspot areas of COD (2006-2012).

	p-value	season	year	sampling site
hotspot	0.001	summer	2012	Purmoj.10k
		fall	2012	Purmoj.10k
secondary	0.001	summer	2011	Purmoj.10k
		fall	2011	Purmoj.10k

Table 3.3 The Hotspot areas of turbidity (2006-2012).

	p-value	season	year	sampling site
hotspot	0.001	spring	2006	Purmoj.10k
		spring	2006	Lapua 9900
		spring	2006	Vt8Oravainen
		spring	2006	Koivulahti vt8
		spring	2006	Maunula
		spring	2006	VaasaKorsnäs
		spring	2006	Harrström a
		spring	2006	Närpiö9200
		spring	2006	Tiukka
		summer	2006	Purmoj.10k
		summer	2006	Vt8Oravainen
		summer	2006	Harrström a
		fall	2006	Lapua 9900
		fall	2006	Vt8Oravainen
		fall	2006	Koivulahti vt8
		fall	2006	VaasaKorsnäs
		fall	2006	Närpiö9200
		spring	2007	Vt8Oravainen
		spring	2007	Maunula
		spring	2007	Harrström a
		spring	2007	Närpiö9200
		spring	2007	Tiukka
		summer	2007	Vt8Oravainen
		summer	2007	Harrström a
		fall	2007	Vt8Oravainen
secondary	0.001	summer	2012	Purmoj.10k

Table 3.4 The Hotspot areas of low pH (2006-2012).

	p-value	season	year	sampling site
hotspot	0.001	fall	2006	Lestj.Kattilakosk
		fall	2006	Perhonj10600
		fall	2006	Kruununpyyn. ala
		fall	2006	Ähtävänj10300
		fall	2006	Purmoj.10k
		fall	2006	Lapua 9900
		fall	2006	Vt8Oravainen
		spring	2007	Kruununpyyn. ala
		spring	2007	Purmoj.10k
		spring	2007	Lapua 9900
		spring	2007	Vt8Oravainen
secondary	0.001	fall	2006	Maunula
		fall	2006	VaasaKorsnäs
		fall	2006	Harrström a
		fall	2006	Närpiö9200
		fall	2006	Tiukka
		spring	2007	Maunula
		spring	2007	Harrström a
		spring	2007	Närpiö9200
		summer	2007	Närpiö9200
		fall	2007	Purmoj.10k
		fall	2007	Lapua 9900
		fall	2007	Vt8Oravainen
		fall	2007	Maunula
		fall	2007	VaasaKorsnäs
		fall	2007	Harrström a
		fall	2007	Närpiö9200
		fall	2007	Tiukka
		spring	2008	Purmoj.10k
		spring	2008	Lapua 9900
		spring	2008	Vt8Oravainen
		spring	2008	Koivulahti vt8
		spring	2008	Maunula
		spring	2008	Harrström a
		spring	2008	Närpiö9200

Many hotspot areas included the year 2006, however diatoms were not surveyed in this year. Therefore, we tried Echelon analysis again excluding year 2006. Each Echelon dendrogram is shown in Figures 3.5, 3.6, 3.7 and 3.8, and Tables 3.5, 3.6, 3.7 and 3.8 explain the hotspot areas.

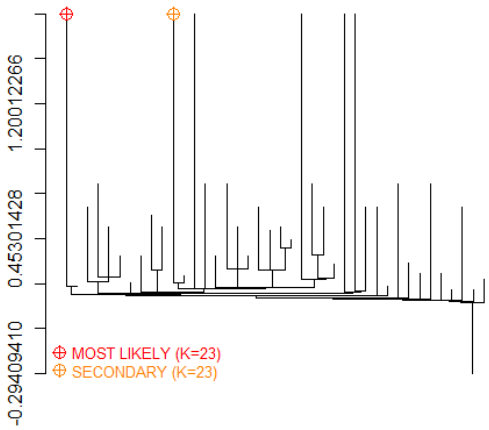


Fig. 3.5 The Echelon dendrogram of alkalinity (2007-2012).

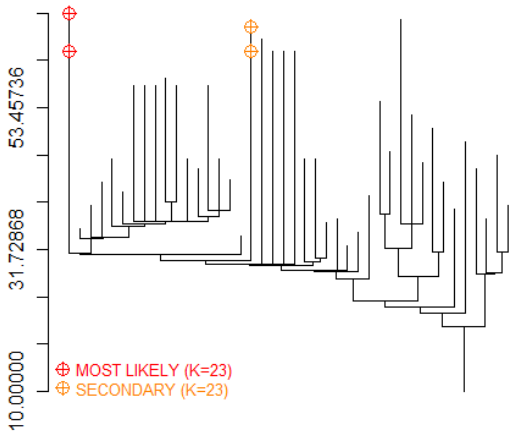


Fig. 3.6 The Echelon dendrogram of COD (2007-2012).

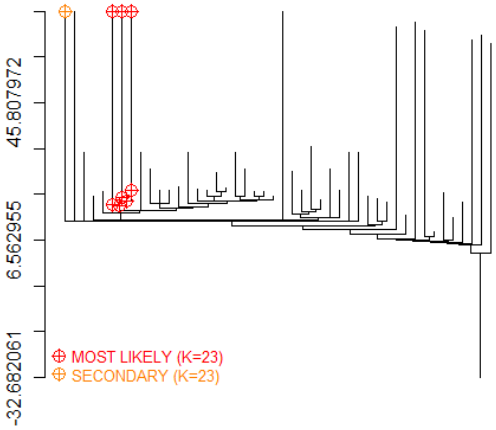


Fig. 3.7 The Echelon dendrogram of turbidity (2007-2012).

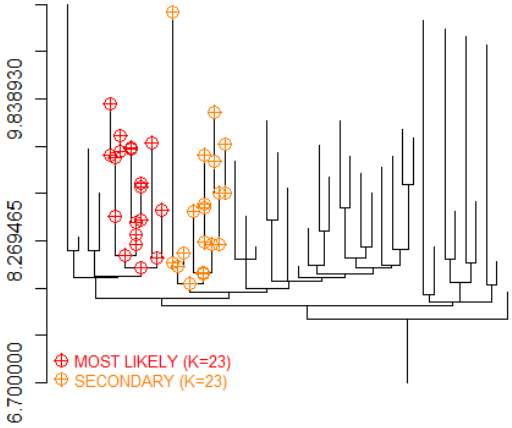


Fig. 3.8 The Echelon dendrogram of low pH (2007-2012).

Table 3.5 The Hotspot areas of alkalinity (2007-2012).

	p-value	season	year	sampling site
hotspot	0.001	summer	2012	Purmoj.10k
secondary	0.001	summer	2011	Purmoj.10k

Table 3.6 The Hotspot areas of COD (2007-2012).

	p-value	season	year	sampling site
hotspot	0.001	summer	2012	Purmoj.10k
		fall	2012	Purmoj.10k
secondary	0.001	summer	2011	Purmoj.10k
		fall	2011	Purmoj.10k

Table 3.7 The Hotspot areas of turbidity (2007-2012).

	p-value	season	year	sampling site
hotspot	0.001	spring	2007	Purmoj.10k
		summer	2007	Purmoj.10k
		fall	2007	Purmoj.10k
		spring	2008	Purmoj.10k
		summer	2008	Purmoj.10k
		fall	2008	Purmoj.10k
		spring	2009	Purmoj.10k
		summer	2009	Purmoj.10k
secondary	0.001	summer	2012	Purmoj.10k

Table 3.8 The Hotspot areas of low pH (2007-2012).

	p-value	season	year	sampling site
hotspot	0.001	spring	2007	Maunula
		spring	2007	Harrström a
		spring	2007	Närpiö9200
		summer	2007	Närpiö9200
		fall	2007	Purmoj.10k
		fall	2007	Lapua 9900
		fall	2007	Vt8Oravainen
		fall	2007	Maunula
		fall	2007	VaasaKorsnäs
		fall	2007	Harrström a
		fall	2007	Närpiö9200
		fall	2007	Tiukka
		spring	2008	Purmoj.10k
		spring	2008	Lapua 9900
		spring	2008	Vt8Oravainen
		spring	2008	Koivulahti vt8
		spring	2008	Maunula
		spring	2008	Harrström a
		spring	2008	Närpiö9200
secondary	0.001	fall	2008	Lestj.Kattilakosk
		fall	2008	Perhonj10600
		fall	2008	Purmoj.10k
		fall	2008	Lapua 9900
		fall	2008	Vt8Oravainen
		fall	2008	Koivulahti vt8
		fall	2008	Maunula
		fall	2008	VaasaKorsnäs
		fall	2008	Harrström a
		fall	2008	Närpiö9200
		fall	2008	Tiukka
		spring	2009	Perhonj10600
		spring	2009	Kruununpyyn. ala
		spring	2009	Ähtävänj10300
		spring	2009	Purmoj.10k
		spring	2009	Lapua 9900
		spring	2009	Vt8Oravainen

	spring	2009	Koivulahti vt8
	spring	2009	Maunula
	spring	2009	Harrström a
	spring	2009	Närpiö9200
	summer	2009	Kruununpyyn. ala

3.2 Water quality impact on diatom diversity

We assessed the impact of water quality on diatom diversity utilizing CART (Figure 3.9 and 3.10). We got different results on diatom indexes. In case of TT_{40}, the important variables were turbidity, alkalinity and pH. On the other hand, the important substances for PMA were pH and COD.

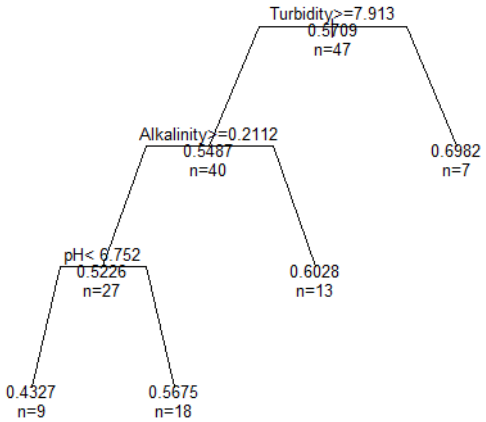


Fig. 3.9 Response variable: TT_{40}.

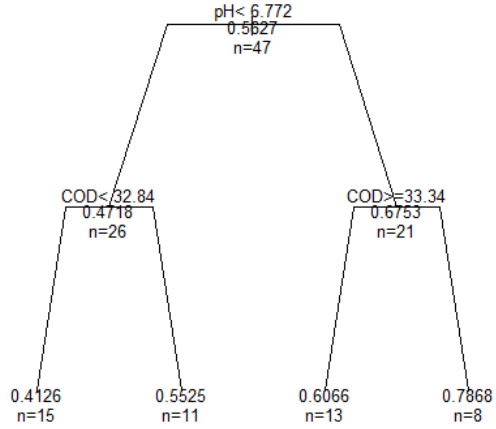


Fig. 3.10 Response variable: PMA.

4. Conclusions

In this study utilizing CART, it was found that the diatom diversity was impacted by multiple variables. On the other hand, we detected statistically significant areas of these variables using Echelon analysis. Water alkalinity and COD did not have a significant effect on the diatom indices. Instead, pH and turbidity had an effect on some sampling sites. However these areas did not match diatom survey area because of the difference of survey time. If the diatom had been surveyed at the same time and the same site of detected hotspot areas, we might have obtained more reliable results. However, we can also assume that the circumstances have not changed between the years. Further studies are needed to explore the role of the catchment area use on the diatom responses.

References

[1] E. Bellier, P. Monestiez, G. Certain, J. Chadœuf and V. Bretagnolle, "Reducing the uncertainty of wildlife population abundance: model-based versus design-based estimates", *Environmetrics*, **24**, pp. 476-488. 2013.

[2] F. Ishioka, K, Kurihara, H. Suito, Y. Horikawa and Y. Ono, "Detection of hotspots for three-dimensional spatial data and its application to environmental pollution data", *Journal of Environmental Science for Sustainable Society*, **1**, pp. 15–24, 2007.

[3] J. Aroviita, E. Koskenniemi, J. Kotanen and H. Hämäläinen, "A priori typology-based prediction of benthic macroinvertebrate fauna for ecological classification of rivers", *Environmental Management*, **42**, pp. 894–906, 2008.

[4] K. Takahashi, T. Yokoyama and T. Tango, "An introduction to disease mapping and disease clustering", *Journal of the National Institute of Public Health*, **57**, 2, pp. 86-92, 2008.

[5] M. A. Novak and E. W. Bode, "Percent model affinity: a new measure of macroinvertebrate community composition", *Journal of North American Benthological Society*, **11**, pp. 80–85, 1992.

[6] M. B. LaBeau, D. M. Robertson, A. S. Mayer, B. C. Pijanowski and D. A. Saad, "Effects of future urban and biofuel crop expansions on the riverine export of phosphorus to the Laurentian Great Lakes", *Ecological Modelling*, **277**, pp. 27-37, 2014.

[7] W. Myers, G. P. Patil and K. Joly, "Echelon approach to areas of concern in synoptic regional monitoring", *Environmental and Ecological Statistics*, **4**, pp. 131–152, 1997.

[8] Wei-Yin Loh, "Classification and regression trees", *WIREs Data Mining and Knowledge Discovery*, **1**, pp. 14-23, 2011.

WP 5

Dissemination of the project results

Aquatic modelling in Finland – experiences and visions

Special lecture at Okayama University
13.5.2014
Timo Huttula and Janne Ropponen

SYKE

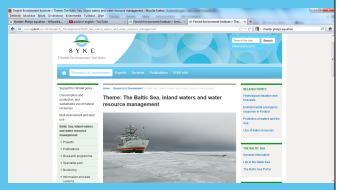
Finland

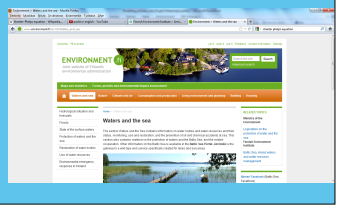
- Total area: 338,144 km² square km 's, of which 10% is water and 69% forest; Europe's largest archipelago, including the semi-autonomous province of Åland.
- Distances: 1,160 km north to south, 540 km west to east Finland's land border with Russia (1,269 km) is the eastern border of the European Union.
- Climate: marked by cold winters and fairly warm summers. Temperatures of −20 Celsius are not uncommon in many areas. Finnish Lapland invariably has the lowest winter temperatures
- Population: 5.4 million, 71% live in towns or urban areas, 29% in rural areas
- GNP: In 2007, Finland's GNP per capita was 34,003 euros

SYKE

SYKE?

- One of 18 state research institutes http://www.research.fi/en/research_environments/state_research_institutes
- Web page www.syke.fi and www.environment.fi
- University of Jyvaskyla www.jyu.fi





Catchlake-hanke

17.5.2010

SYKE

Why water quality models

- New legislation in late 1960's forced point loaders (like industry and municipalities) to apply permission to lead purified waste waters to recipient waters (lakes, rivers, coastal areas etc..) from the Water Courts and
- During the Assesment Procedure ("Katselmustoimitus") the effects on the water quality was assessed and also the compensations were ordered
- Models were applied for
  - Determining the optimal loads
  - Determining the effected geographical areas
  - Dtermining the optimal locations for water water outlets
- Model applications from 1970's by Virtanen, Koponen and Sarkkula

SYKE

Later needs for modelling

- Problems with eutrophication and related diffuse loads increased the need of WQ assessment and modelling
- Effects of global change need to be assessed
- European wide approach: Water Framework Directive
- Fate and transport of harmfull substances (heavy metals, organic compounds etc..) is to be predicted since many recipient waters are also used for water supply
- Hydrological (water resources) modelling was fist needed for optimal regulation of our large lakes. Modelling started in early 1980's by Vehviläinen

SYKE

Presently many new needs

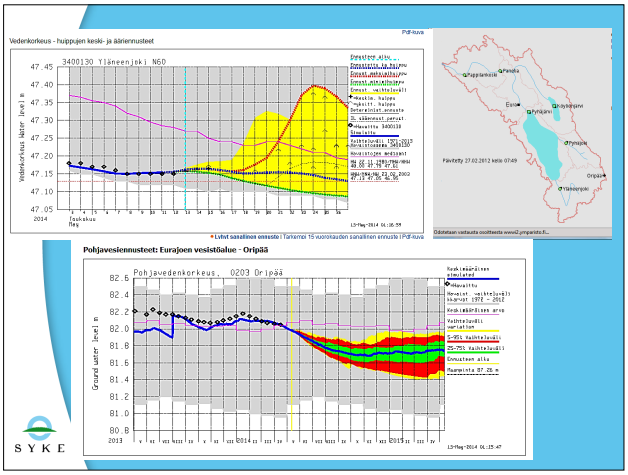
- Oil spill models
  - Operative models for oil combatting
- Habitat modelling for river restoration
- Socio-economical modelling
  - See CONPAT-project

SYKE

1. Catchment models

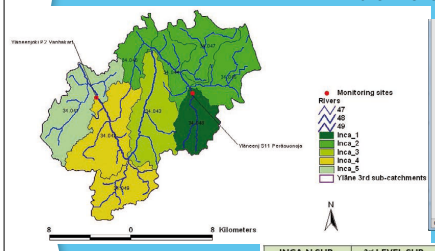
Hydrological simulation and forecast

- A country wide system (WSFS):
  - http://www.ymparisto.fi/en-US/Waters_and_sea/Hydrological_situation_and_forecasts/Hydrological_forecasts_and_maps/Hydrological_forecasts_and_maps%2823179929
  - Used for operational forecasts
  - Used also for research
  - Based on water balance computation on watersheds (3rd division level)
- Dynamic links
  - Meteorological data forecasts (FMI)
  - Hydrological observations (water levels, river discharges, water temperature, ground water level)
  - Links to satellite automated imaging products (snow coverage, snow water equivalent)
- Automatic and tailored products



INCA-N

- process-based and semi-distributed model that integrates hydrology, catchment and river N processes to simulate daily concentrations of NO_3-N and NH_4-N in the river system

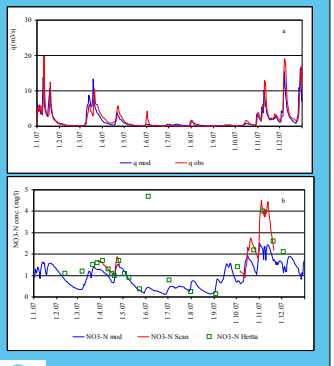


River Yläneenjoki catchement delineation for INCA-N application (Lepistö *et al.* 2008)

INCA-N SUB-CATCHMENTS	3rd LEVEL SUB-CATCHMENTS	AREA (Km²)	AGRICULTURAL AREA (%)
1	34.048	20.08	28
2	34.044-34.047	78.18	42
3	34.043	37.09	26
4	34.042, 34.049	61.59	27
5	34.041	36.10	19

INCA-N

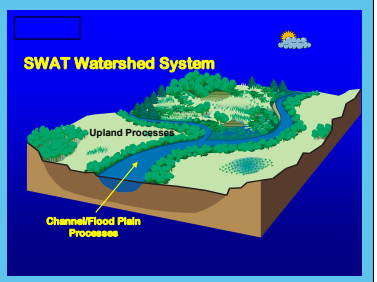
- Calibration was successful The model was able to simulate annual dynamics of discharge and flow
- The NO_3 -N concentration peaks were sometimes slightly under- or overestimated
- No bigger problems were reported in CatchLake project report



INCA-N results for year 2007 in Vanhakartano: (a) modelled vs. observed discharge, (b) modelled vs. observed NO_3-N concentration. (Lepistö *et al.* 2008)

SWAT

- Continuous time model, operates on a daily time step at catchment scale
- Simulates suspended sediment & nutrient loading on catchment scale
- In European wide use
- Collaboration with Okayama University (Lake Kojima catchment)
- Has potential to include agricultural management actions



(See more in the the project report: http://www.ymparisto.fi/download.asp?contentid=85389&lan=fi)

SWAT

- According to CatchLake project report, the parameterization was again found challenging
- Sensitivity analysis and auto calibration tool were helpful
- Continuous turbidity data showed that model was not able to catch all the relevant erosion processes

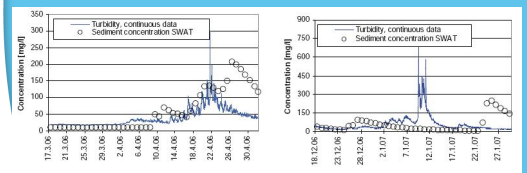


Fig. 27. Simulated sediment concentration and continuously measured turbidity at Vanhakartano during spring 2006 and winter 2007.

Lepistö A. et al. 2008

2. Simple mass balance models for lakes

Lake mass balance

$$\frac{dm}{dt} = I - O - S$$

- Where, m= total amount of phosphorus in a lake, I=total load of phosphorus entering to the lake, O=amount of phosphorus leaving the lake, S=retention (the amount of phosphorus (P) sediment to the lake)
- First applications by Piontelli and Tonolli (1969)
- In Vollenweiders (1969) solution the lake was considered as a continuously stirred tank (CSTR) and sedimentation of P was taken as following the first order kinetics with settling coefficient σ and phosphorus concentration (c) at the outlet was taken as mean concentration in the lake

Vollenweiders equation

$$\frac{dm}{dt} = I - Qc - \sigma m$$

$$\rho = \frac{Q}{V} \Rightarrow \frac{dc}{dt} = \frac{I}{V} - \rho c - \sigma c$$

- Where, Q= water discharge at the outlet and ρ= exchange coefficient, σ=settling coefficient
- It is expected here that the lake volume is constant (eq. an annual mean) and introduce
- In steady state we mark phosphorus concentration as c_{ss} it can be solved for time t

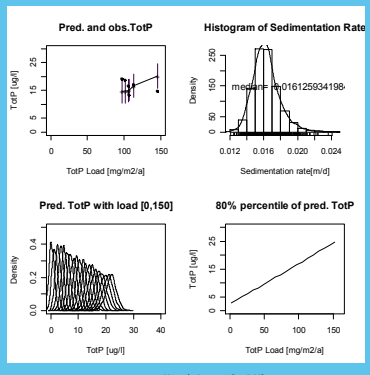
Solution for Vollenweider-equation

$$c = c_{ss} - (c_{ss} - c_0)e^{-(\rho+\sigma)(t-t_0)}$$

- Where c_0 is phosphorus concentration at time t_0
- Comments
  - Settling description with first order kinetics is too simple
  - Model can be used for studying the effects of loading options

ToTP forecasts

Järven ravinnepitoisuus vs. ravinnekuormitus . Lasketut ja havaitut arvot



Estimoitu sedimentationopeuden jakauma

TotP-jakauman ennusteet kuormituksilla [0,10,20, ... ,150] kg/d

TotP-ennusteiden 80% fraktiili (20% ylitystodennäköisyys) kuormituksen funktiona

3. River models

First models for flowing waters

$$\frac{dC}{dt} = K_2(C_s - C) = K_2 D$$

$$-\frac{dL}{dt} = K_1 L$$

$$C = C_s - \frac{K_1 L_0}{K_2 - K_1}((e^{-K_1 t} - e^{-K_2 t}) + (C_s - C_0)e^{-K_2 t})$$

- Oxygen model by Streeter-Phelps (1925)
- Oxygen = f(advection by river waters, aeration and decomposition by bacteria), steady state
- In equation:
  - C=oxygen concentration (mg/l), C_0 =initial oxygen concentration in waters, C_s=saturation concentration of oxygen, D=oxygen deficit, K_2 = aeration coefficient (like 0,15 1/d), L=BOD value (mg/l), L_0= BOD- initial value , K_1 = decay coefficient of BOD (like 0,25 1/d), T=time (days)

Using Streeter-Phelps

- Water temperature and oxygen saturation concentration in that temperature have to be known.
- K_1 is temperature dependent. K_2 only slightly
- Several temperature correction equations available. Frisk and Nyholm (1980) mostly used in Finland
- Flow time has to be calculated for each river reach. In this way advection is taken into account

Dynamic flow models for rivers

$$Q = \frac{WY^{3/5}}{n}\left(S_0 - \frac{\partial Y}{\partial x} - \frac{V}{g}\frac{\partial V}{\partial x} - \frac{1}{g}\frac{\partial V}{\partial t}\right)^{1/2}$$

1 2 3 4 5

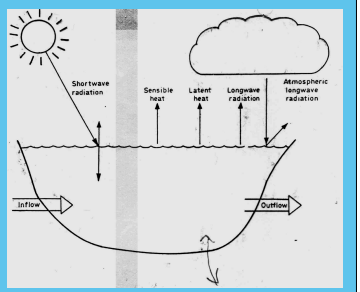
- St. Venant (1848) for river dynamics
- In equation
  - 1=discharge, 2=pressure gradient=bottom sloping term, 3=pressure gradient due to the surface slope, 4=advection due to the river flow, 5=local acceleration term
- For WQ -simulations
  - Streeter-Phelps type equation
  - Suspended solids like we look later in 3D models
  - Nutrients also similarly
- Presently we use SOBEK model from Deltares and HEC-RAS from US/EPA

4. Water temperature models

Heat balance components

- Components can be solved by quite well known equations
- Snow and ice form special challenges in Northern conditions

Sometimes also heat from precipitation and sediment should be calculated



Equations in PROBE-model, F_s

$$F_s = (1-a)S_0 \cos z(T_r - A_w)\prod_{i=1}^{3}(1-N_i(-T_i))$$

$$A_w = 0.077(u \sec z)^{0.3}$$

$$T_r = 1.041 - 0.16(\sec z)^{0.5}$$

$$T_{low} = 0.35 - 0.015 \sec z$$

$$T_{middle} = 0.45 - 0.01 \sec z$$

$$T_{high} = 0.9 - 0.04 \sec z$$

$$\sec z = (\cos z)^{-1}$$

a=albedo

A_w= absorption by the water vapor

N_i= Ni is the amount of clouds of the different categories

(low, middle, high).

S_0=solar constant, 1395 Wm^{-2}

T_r= scattering-transmission function= f(z)

T_i= cloud function

u= amount water in the air mass

z=solar zenit angle

Equations in PROBE-model, Long wave radiation F_l

$$F_l = F_l \uparrow - F_l \downarrow$$

$$F_l \uparrow = \varepsilon' \sigma T_s^4$$

$$F_l \downarrow = \sigma T_a^4 (c + b\sqrt{e_a})(1 + dN)$$

ε'= emissivity of the lake water =0.97

σ=Stefan-Bolzman constant=$5.67*10^{-8}$ Wm^{-2} K^{-4}

T_s=lake water surface temperature (^{0}K)

T_a= air temperature (^{0}K)

e_a= water vapour pressure in air (mb)

N= cloudiness

c,b ja d = constant

Equations in PROBE-model Sensitive heat flux, F_c

$$F_c \downarrow\uparrow = \rho_a C_p \bar{U}(C_{c1} - C_{c2}(T_s - T_a))$$

$$S_t = \bar{U}(T_s - T_a)$$

ρ=air temperature, kg m^{-3} Cp=specific heat of water =4200 J kg^{-1} ,U= wind velocity m s^{-1} S_t= air stability, C_{c1}, C_{c2} = are sensible heat transfer, which depend on air stability

In stable conditions (S_t<0) C_{c1} = 0.0026 and C_{c2} =0.86E^{-3}

In unstable conditions (0<S_t<25), C_{c1} = 0.002 and C_{c2} =0.97E^{-3}

Very unstable conditions (25<S_t), C_{c1} = 0.0 and C_{c2} =1.46E^{-3}

Equations in PROBE-model...

Latent heat flux, F_e

$$F_l \downarrow = LC_e \bar{U}(Q_w - Q_a)$$

L=L is the latent heat of evaporation, Ce, is the moisture transfer coefficient and Qw and Qa are the water vapour densities close to the water surface and in the atmosphere respectively.

Total heat flux

$$F_N = F_l + F_c + F_e$$

PROBE: Heat equation and vertical mixing

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(\frac{\nu_T}{\sigma_T}\frac{\partial T}{\partial z}\right) + S_T$$

$$\nu_T = \frac{C_v \rho k^2}{\varepsilon}$$

Boundary condition at the upper boundary

$$\frac{\nu_T \partial T}{\sigma_T \partial z} = \frac{F_N}{\rho C_p}$$

ν_t= eddy diffusivity

σ_t=turbulent Prandtl number= ν/γ

ν=kinematic viscosity

γ=heat conductivity

k=kinetic turbulent energy

ε=dissipation of turbulent energy

C_v=empirical constant

Experiences about PROBE-model

- Case study: Huttula ym. 1994 Effects of Climate Change....pdf (see reference list below)
- Good results in scales from days...to years
- Water balance well calculated
- Ice formation and decay well calculated
- Heat exchange coefficients need to be calibrated for some lakes
- Hypolimnetic temperatures too low sometimes ← vertical mixing too small in model
- Sheltering effects, effects of sediment quality and penetration of short wave radiation (extinction coefficients) need special attention

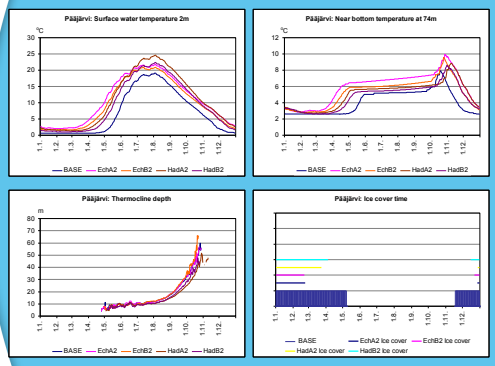
MyLake

- One dimensional vertical lake model
- Andersson and Saloranta 2000
- Ice model by Leppäranta (1991) and Saloranta (2000)
- The vertical diffusion coefficient from the stability frequency N^2 (Hondzo and Stefan, 1993)
- Utilises the MATLAB *Air-Sea Toolbox* (http://sea-mat.whoi.edu/air_sea-html/) for calclulation of radiative and turbulent heat fluxes, surface wind stress and astronomical variables
- Vertical mixing is based on the energy calculation between kinetic energy from wind and potential energy of layer(s) to be mixed
- **Model package with documentation available from TH**

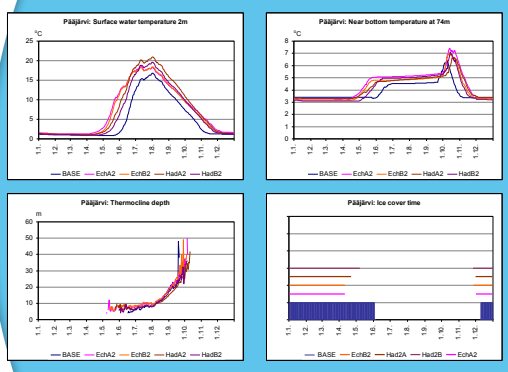
FINESSI-project

- Web tool for assessing the effects of global chage in Finland www.finessi.info/finessi
- Lake Pääjärvi (area= 13.5 km^2 , max depth= 87 m) and Halsjärvi (area= 0.5 km^2 , max depth= 6 m)
- Meteorological data from Jokioinen
- Calibration for Pääjärvi Link
- FINESSI uses 6 GCM-models Link
  - We Echam and Hadley Center

Jokioinen



Sodankylä



5. Modelling currents in lakes and seas

Basic laws

- Conservation of momentum with Boussinesq approximation* and the assumption of vertical hydrostatic equilibrium**
- Continuity equation
- Conservation energy
- Equation of state

$$\frac{\partial u}{\partial t} + (u \cdot \nabla)u + 2\Omega \times u = -g + \rho^{-1}\nabla p + \nu\nabla^2 u + A_H \nabla_H{}^2 u + A_V \partial^2 u / \partial z^2$$

adv. rotation gravity press. int.frict. turbulenceterms

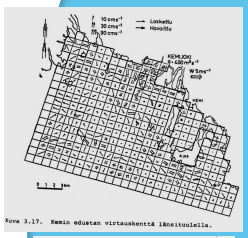
$$\nabla \cdot \bar{u} = 0 \Leftrightarrow \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \qquad \rho = \rho_0 + \alpha(T - T_0) + \beta(S - S_0)$$

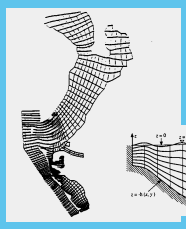
* It states that density differences are sufficiently small to be neglected, except where they appear in terms multiplied by *g*, the acceleration due to gravity

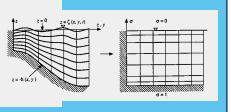
** The principle of hydrostatic equilibrium is that the pressure at any point in a fluid at rest is just due to the weight of the overlying fluid

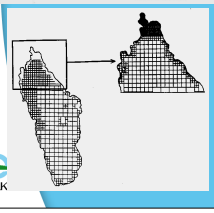
Boundary conditions

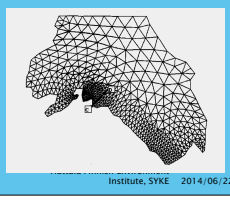
- Boundary conditions
  - Surface: wind, solar radiation, net longwave heat balance ← meteorological data (IL)
  - Sides: river discharges, concentrations of ← hydrological and water quality data (SYKE)
  - Bottom: friction …← calibrated using bathymetric and vegation maps
  - Can be given
    - As fixed values, velocity=0 at bottom
    - As fluxes, sediment heat flux, about 1-3 W/m2
    - As sliding conditions (frictionless boundary), velocity on the side same as in the lake









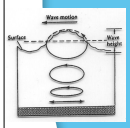


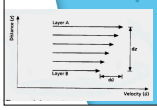
Sedimentation of particulate matter

$$v_f = \frac{2(\rho_s - \rho)gr^2}{9\mu} = \frac{(\rho_s - \rho)gd^2}{18\mu}$$

- Settling velocity is solved from Stokes equation with some correction factor
- Material flux at both due to the erosive forces are calculated by using crtical shear approach
- Total shear (τ) on the lake bottom =
  - shear by orbital movements of waves = f(wind fetch over lake, lake mean depth, wind velocity and duration)
  - shear by currents
  - τ > critical shear (τ_{cr}), erosion happens with a rate $\propto a*(\text{excess shear})^b$
  - τ_{cr}, a and b are experimental values, which we calibrate during model application

values for τ_{cr}: 0.008…1 Nm^{-2}, b=1..3, a = depends on sediment





6. Modelling water quality or hydrochemical and hydrobiological modelling

Application of WQ-models

- We include :
  - Advection
  - Dispersion
  - Settling on the bottom
- Bio- chemical processes
  - Decomposition, respiration, aeration, anaerobic release of P from the bottom
  - Select the most important variables concerning the problem
  - Oxygen, nutrients (like P,N), chlorophyll-a and some conservative substance (like Na)
  - Limiting factors (light, nutrients, …) must be included. Check!
  - Temperature corrections must be included. Check!

Concentration equation

$$\frac{\partial c}{\partial t} = \frac{\partial qL}{\partial n} - u\frac{\partial c}{\partial x} - v\frac{\partial c}{\partial y} - w\frac{\partial c}{\partial z} + \frac{\partial c}{\partial x}(D_x\frac{\partial c}{\partial x}) + \frac{\partial c}{\partial y}(D_y\frac{\partial c}{\partial y}) + \frac{\partial c}{\partial z}(D_z\frac{\partial c}{\partial z}) + R(T,c...)$$

where,

- c = concentration of substance (nutrient, oxygen, metals, algae..) , qL= amount of loading release , n= length measure against release, u,v,w = advective velocities in x–, y– ja z– directions, D_x, D_y, D_z = dispersion coefficients, R(T,c) = biogeochemical changes affecting to the concentration of the substance
- Selection of the substance depends on the problem to be solved and available data

Oxygen model (Malve, Huttula and Lehtinen 1992)

Computation of dissolved oxygen Both abiotic and biotic factors affect the concentration of oxygen. The change of dissolved oxygen concentration as a function of time is described by the following equation':

$$\frac{dO_2}{dt} = K' \times \sqrt{W} \times (O_{2sat} - O_2) - K_1 \times BOD_7 \times BRAT + \mu \times \alpha_1 \times CH - r \times \alpha_2 \times CH - \frac{SOD \times AREA}{V} \quad (7)$$

K' = aeration constant = 2.0 • 10^{-4} cm/d, W = wind speed, z = layer thickness
O_{2sat} = dissolved oxygen saturation concentration at the surface layer temperature
O_2 = dissolved oxygen concentration at the surface layer temperature
K_1 = BOD decay rate = 0.1 1/d (function of temperature, f(T))
BOD_7 = BOD_7 concentration
BRAT = BOD/BOD_7 = 1.5
α_1, α_2 = stoichiometric coefficients for growth and respiration= 0.1903
μ = growth rate of algae
r = algal respiration coefficient = 0.065 1/d, (f(T))
CH = chlorophyll concentration
SOD = bottom sediment oxygen demand, (f(T))
AREA = area of the bottom sediment
V = volume of the water body

The first term on the right hand side describes aeration in the surface layer, the second one biological oxygen demand, the third and fourth ones phytoplankton growth and respiration, and the last one bottom sediment oxygen demand.

Phytoplankton biomass and ToTP

Computation of phytoplankton biomass Chlorophyll-a concentration is used as a relative measure of phytoplankton biomass. The rate of change of phytoplankton biomass is expressed as [8]

$$\frac{dCH}{dt} = \mu \times CH - r \times CH - \frac{SED}{h} \times CH \quad (8)$$

Phosphorus cycle Description of the phosphorus cycle is quite simple. Total phosphorus concentration in the lake is affected by external loading, phosphorus sedimentation and release of phosphorus under unaerobic conditions.
The change of total phosphorus concentration as a function of time is described by the following equation [8]

$$\frac{dTOTP}{dt} = -\frac{SEDP}{h} \times (TOTP)^2 + LOAD + \frac{RELEASE}{AREA \times h} \quad (11)$$

SEDP = net phosphorus sedimentation coeffkient = 0.002 (m/d)/(µg/l)
LOAD = external loading
RELEASE = rate of phosphorus release from the sediment under anaerobic conditions

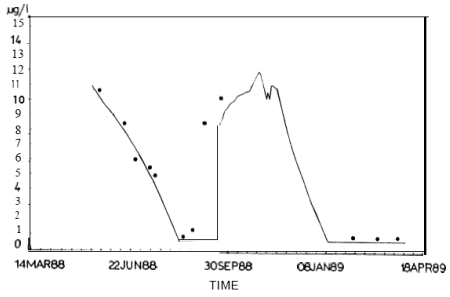


Figure 7. Observed (.) and calculated (–) oxygen concentrations in bottom layer (height 1 m). Calibration, summer 1988.

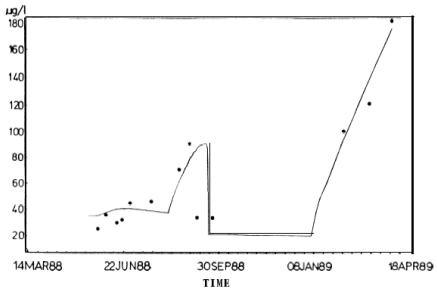


Figure 8. Observed (.) and calculated (–) total phosphorus concentrations in bottom layer (height 1 m). Calibration, summer 1988.

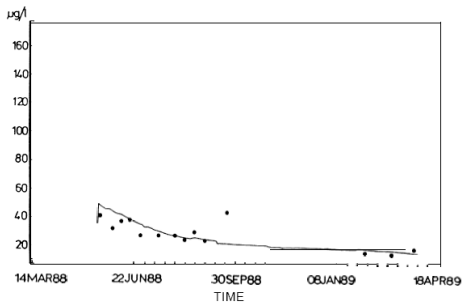


Figure 9. Observed (.) and calculated (–) total phosphorus concentrations in surface (0-1 m) layer. Calibration, summer 1988.

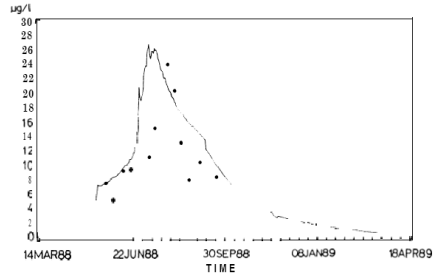


Figure 10. Observed (.) and calculated (–) chlorophyll-a concentrations in surface (O-1 m) layer. Calibration, summer 1988.

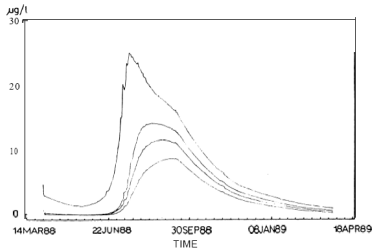


Fig. 11 Calculated chlorophyll-a concentration with loading levels 1 ▪ 4.

1. Present total phosphorus loading 55 697 kg/year

2. Fast obtainable loading level 47 469 kg/year
-reduction of the present loading 14.8%

3. Desirable loading level 37 483 kg/year
-reduction of the present loading 32.7%

4. Loading level with the best available protection measures 30 689 kg/year
-reduction of the present loading 44.9%

Interactions in EIA-SYKE-model

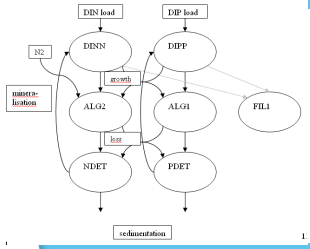


Table 1. Model variables.

symbol	definition	unit
c_C	Biomass of cyanobacteria (wet weight)	g m^{-2}
c_A	Biomass of the other algae (wet weight)	g m^{-2}
c_{DIN}	DIN concentration	mg m^{-3}
c_{DIP}	DIP concentration	mg m^{-3}
c_{Ndet}	Detritus nitrogen	mg m^{-3}
c_{Pdet}	Detritus phosphorus	mg m^{-3}
L_{DIN}	DIN load	mg m^{-3} d^{-1}
L_{DIP}	DIP load	mg m^{-3} d^{-1}
I	Total radiation	MJ m^{-2}d^{-1}
T	Temperature	^{o}C
Ice	Ice-cover (0,1)	-
t	Time	d
h	Depth of grid cell	m

Table 2. Model parameters

Symbol	definition	reference	value	unit
N_{inC}	Nitrogen in cyanobacteria	Redfield, 1958	0.0193	-
P_{inC}	Phosphorus in cyanobacteria	Redfield, 1958	0.00268	-
N_{inA}	Nitrogen in the other algae	Redfield, 1958	0.0193	-
P_{inA}	Phosphorus in the other algae	Redfield, 1958	0.00268	-
μ_{Cmax}	Maximal growth rate of cyanobacteria	calibration	0.5	d^{-1}
μ_{Amax}	Maximal growth rate of the other algae	Olli *et al.*, 1996	0.7	d^{-1}
R_{Cmax}	Maximum loss rate of cyanobacteria	calibration	0.1	d^{-1}
R_{Amax}	Maximum loss rate of the other algae	calibration	0.15	d^{-1}
K_{NC}	Half-saturation coefficient of DIN for cyanobacteria	Tyrrell, 1999	0	mg m^{-3}
K_{PC}	Half-saturation coefficient of DIP for cyanobacteria	Kononen & Leppänen, 1997	2	mg m^{-3}
K_{NA}	Half-saturation coefficient of DIN for the other algae	calibration	7	mg m^{-3}
K_{PA}	Half-saturation coefficient of DIP for the other algae	calibration	1	mg m^{-3}
K_{IC}	Half saturation coefficient of radiation for cyanobacteria	calibration	20	MJ m^{-2} d^{-1}
K_{IA}	Half saturation coefficient of radiation for the other algae	calibration	15	MJ m^{-2} d^{-1}
C_{min}	Minimum biomass of cyanobacteria	calibration	0.5	g m^{-2}
A_{min}	Minimum biomass of the other algae	calibration	0.01	g m^{-2}
A_{max}	Maximum total biomass of algae	calibration	300	g m^{-2}
β_0	Maximal detritus nitrogen mineralisation rate	Garber, 1984	0.018	d^{-1}
γ_0	Maximal detritus phosphorus mineralisation rate	Garber, 1984	0.043	d^{-1}
v_{NDet}	Settling rate of detritus nitrogen	Heiskanen & Tallberg, 1999	1	m d^{-1}

v_{NDet}	Settling rate of detritus nitrogen	Heiskanen & Tallberg, 1999	1	m d^{-1}
v_{PDet}	Settling rate of detritus phosphorus	Heiskanen & Tallberg, 1999	1	m d^{-1}
S_{NDet}	Sedimentation rate of detritus nitrogen	calibration	0.16	m d^{-1}
S_{PDet}	Sedimentation rate of detritus phosphorus	Lehtoranta, 1998	0.00	m d^{-1}
T_{opt}	Optimal temperature			
	for the growth of cyanobacteria	Kononen & Leppänen, 1997	25	^{o}C
	for the growth of the other algae	calibration	15	^{o}C
	for losses	calibration	25	^{o}C
	for detritus nitrogen mineralisation	Garber, 1984	18	^{o}C
	for detritus phosphorus mineralisation	Garber, 1984	18	^{o}C
a_T	Coefficient for temperature limiting factor			
	for the growth of cyanobacteria	calibration	1.14	-
	for the growth of the other algae	calibration	1.001	-
	for losses	calibration	1.05	-
	for detritus nitrogen mineralisation	Garber, 1984	1.31	-
	for detritus phosphorus mineralisation	Garber, 1984	1.60	-
I_{red}	Radiation attenuation by ice	calibration	0.5	-
h_{mix}	Depth of mixing layer	calibration	20	m

Table 3. Model equations, rates and limiting factors.

Equations

$$\frac{\partial c_C}{\partial t} = (\mu_C - R_C)c_C \quad (1)$$

$$\frac{\partial c_A}{\partial t} = (\mu_A - R_A)c_A \quad (2)$$

$$\frac{\partial c_{DIN}}{\partial t} = \beta\, c_{NDet} - \mu_A N_{inA} c_A h_{mix}^{-1} - \mu_C N_{inC} c_C h_{mix}^{-1} + L_{DIN} \quad (3)$$

$$\frac{\partial c_{DIP}}{\partial t} = \gamma\, c_{PDet} - \mu_A P_{inA} c_A h_{mix}^{-1} - \mu_C P_{inC} c_C h_{mix}^{-1} + L_{DIP} \quad (4)$$

$$\frac{\partial c_{NDet}}{\partial t} = N_{inA}\ R_A c_A h_{mix}^{-1} + N_{inC}\ R_C c_C h_{mix}^{-1} - \beta c_{NDet} - v_{NDet} c_{NDet} h^{-1} - S_{NDet} c_{NDet} h^{-1} \quad (5)$$

$$\frac{\partial c_{PDet}}{\partial t} = P_{inA}\ R_A c_A h_{mix}^{-1} + P_{inC}\ R_C c_C h_{mix}^{-1} - \gamma c_{PDet} - v_{PDet} c_{PDet} h^{-1} - S_{PDet} c_{PDet} h^{-1} \quad (6)$$

Rates

$$\mu_C = \mu_{Cmax} f_{CN}(c_{DIN}, c_{DIP})\ f_{CI}(I)\ f(T) f_{AC}(c_A, c_C) \quad (7)$$

$$\mu_A = \mu_{Amax} f_{AN}(c_{DIN}, c_{DIP})\ f_{AI}(I)\ f(T) f_{AC}(c_A, c_C) \quad (8)$$

$$R_C = R_{Cmax} f(T)(c_C - C_{min})/c_C \quad (9)$$

$$R_A = R_{Amax} f(T)(c_A - A_{min})/c_A \quad (10)$$

$$\beta = \beta_0 f(T) \quad (11)$$

$$\gamma = \gamma_0 f(T) \quad (12)$$

Limiting factors

$$f_{DIN}(c_{DIN}, c_{DIP}) = \frac{c_{DIN}}{c_{DIN} + K_{NC}} \frac{c_{DIP}}{c_{DIP} + K_{PC}} \quad (13)$$

$$f_{AN}(c_{DIN}, c_{DIP}) = \frac{c_{DIN}}{c_{DIN} + K_{NA}} \frac{c_{DIP}}{c_{DIP} + K_{PA}} \quad (14)$$

$$f(T) = \exp\left[\int_{T_{opt}}^{T} \ln \theta dT\right], \text{ where } \theta = a_T + (1 - a_T) T / T_{opt} \quad (15)$$

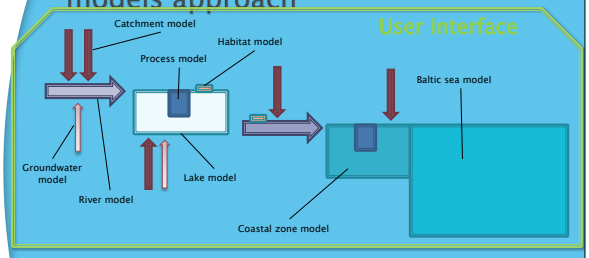
$$f_{CI}(I) = \frac{I(1 - Icel_{red})}{I(1 - Icel_{red}) + K_{IC}} \quad (16)$$

$$f_{AI}(I) = \frac{I(1 - Icel_{red})}{I(1 - Icel_{red}) + K_{IA}} \quad (17)$$

$$f_{AC}(c_A, c_C) = 1 - \frac{c_A + c_C}{A_{max}} \quad (18)$$

7. Some visions and lessons

Modelling philosophy: the chain of models approach

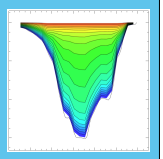


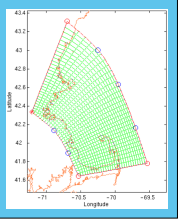
- Water flows through the different models and tracers flow along with the water. Process models can act on the tracers, or calculate small scale flow, and e.g. habitat models depend on the other models
- **Catchment model**: WSFS-Vemala, INCA, SWAT, SOBEK
- **River model**: SOBEK, River2D, COHERENS
- **Habitat model**: Delft HABITAT, River2D
- **Lake model**: COHERENS, MyLake
- **Process models**: biological models, sediment models, Elmer
- **Coastal zone model**: COHERENS
- **Baltic sea model**: COHERENS

COHERENS modelling tool

museum

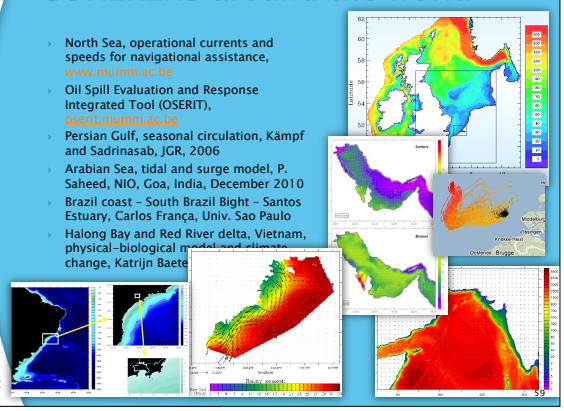
- **CO**upled **H**ydrodynamical **E**cological model for **RE**gio**N**al **S**helf seas, RBINS-MUMM, Belgium
- 3D finite difference, s-layer, multi-purpose numerical model designed for application in coastal and shelf seas, estuaries, lakes and reservoirs
- Open source, available to the public since 2000
- Multi-platform, extremely good documentation (1500+ pages)
- Modular design
  - Physical core
    - selectable simulation modes, dimensions, numerical schemes
  - Flexible and expandable
    - Biological/ecological module
    - Sedimentation module
    - Wave modules
    - Tracers
    - Processes
- Actively developed , constantly evolving (latest version V2.5.1 available since August/2013)





COHERENS around the world

- North Sea, operational currents and speeds for navigational assistance, www.mumm.ac.be
- Oil Spill Evaluation and Response Integrated Tool (OSERIT), oserit.mumm.ac.be
- Persian Gulf, seasonal circulation, Kämpf and Sadrinasab, JGR, 2006
- Arabian Sea, tidal and surge model, P. Saheed, NIO, Goa, India, December 2010
- Brazil coast – South Brazil Bight – Santos Estuary, Carlos França, Univ. Sao Paulo
- Halong Bay and Red River delta, Vietnam, physical-biological m… change, Katrijn Baete

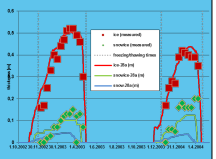


COHERENS at SYKE

- COHERENS was selected as SYKE's marine and lake modelling tool of choice
  - Performance as multi-purpose modelling tool – adaptability to both lake and marine environment
  - Open source code & excellent documentation
  - Modularity
  - Active development
- COHERENS performed well in model inter-comparison study in the Gulf of Finland (Myrberg et al, 2010)
- Development resources can be concentrated to improving a single modelling tool
  - The goal of module development work at SYKE is to improve the applicability of the COHERENS model in low-salinity Baltic region applications for multi-year (incl. winter) simulations
  - Modules have been developed for eg. tracers, sediments and ice formation/melting
- V1 code used since 2006, from 2012 all new projects use V2

Ongoing, near-future and proposed COHERENS developments at SYKE 1/4

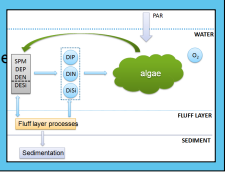
- Adapting model heat balance to boreal lake environment
  - Shallow lake thermodynamics not completely correctly modelled
- Ice and snow model
  - Important for multi-year simulations and climate change scenarios
  - Measurements
  - Calibrate and validate the developed code
  - Add simple advection
- Fully utilise the advances in COHERENS V2 code
  - Nesting
  - Sediment model
  - Flooding and drying schemes
  - etc
- Coupling COHERENS to other models
  - River models, catchment models, WSFS-VEMALA, groundwater flows
  - "Chain of models"



SYKE

Ongoing, near-future and proposed COHERENS developments 2/4

- Increase understanding of shallow area flows
  - Fetch
  - Local wind effects on the same horisontal scale as lake model itself
  - Hydraulic effects of bottom macrophytes
- Wave modelling
  - Coupling to external models
- Ecological/nutrient model with particle-bound nutrient transport
- Transport and evolution of harmful substances
  - Organic tin compounds (TbTs)
  - Bacteria, viruses
  - Artificial sweeteners, medicines
- Coupling to external biological model
  - BFM
  - PyWQM



SYKE

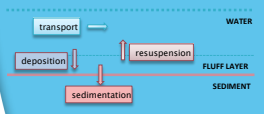
Ongoing, near-future and proposed COHERENS developments 3/4

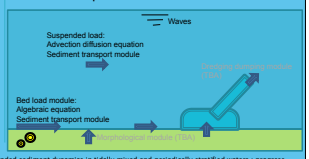
- Age tracers and automatic fraction calculations (2013-2014)
- Automatic sanity checking mid-simulation
- Improving data output usability
  - Looking at results while simulation is running
  - Importing/exporting NetCDF to other software
  - Visualisation improvements
- Seto inland sea, Japan (Okayama U.)
  - Complex morphology
  - Tidal interactions
  - Power generation
- Data assimilation and automatic calibration (collaboration with Okayama U.)
  - Initial values and loads from satellite imagery
  - Particle filtering method
  - Ensemble Kalman filtering

SYKE

COHERENS sediment models

- COHERENS V1
  - Very simple model
  - Single fraction
  - Simple wave-current interaction
  - Temporary deposition in 'fluff layer', resuspension and permanent sedimentation
  - Based on Jones et al. (1995)[1]
  - Area of applicability on ocean floors
- COHERENS V2.5
  - State-of-the-art model
  - Cohesive and non-cohesive sediment
    - Multiple fractions
    - Hiding effects and flocculation
  - Multiple ways to calculate bed-current-wave interactions
  - Two-way coupled with hydrodynamics
    - Density driven sediment flows
    - Turbulence damping
  - Bed slope effects
  - Developed at IMDC & RBINS[2]





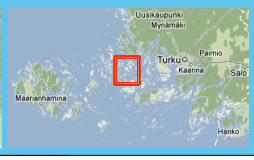
SYKE

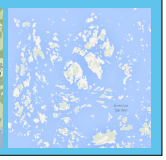
1) Jones S.E., Jago C.F. and Simpson J.H., 1995. Modelling suspended sediment dynamics in tidally mixed and periodically stratified waters : progress and pitfalls, In : C.B. Pattiaratchi (Editor), Mixing Processes in Estuaries and Coastal Seas. Coastal and Estuarine Studies, Vol. 41, American Geophysical Union, 315-338.

2) Breugem, W.A., Decrop, B., Frederix, K., Delecluyse, K., v .Holland, G., Luyten, P., Hyde, P., Development of a sediment transport model in Coherens,

Ongoing, near-future and proposed COHERENS developments 4/4

- Archipelago Sea model
  - Commission from the Ministry of Environment
  - The goal is to produce an intuitive graphical tool to help decision-making in controlling loading to the Archipelago Sea area
  - Nested marine model on the Baltic, coupled to WQ models
    - Full Baltic Sea coarse resolution + fine resolution archipelago sea model
    - River/agricultural loading (WSFS-VEMALA)
    - Eutrophication & water quality (PyWQM/SEABED project)
  - Joint project with SYKE's Freshwater and Marine Centers with partners from FMI, regional authorities, Åbo Akademi, KTH (Sweden)





SYKE

COHERENS applicability

- From large sea areas with high salinity down to small lakes (~ 1 km^2) with zero salinity
- Physical modelling in three dimensions
  - Currents, temperature, salinity
  - Tidal effects
  - Flooding
  - Morphology changes
- Ecological modelling
  - Nutrient fluxes, biology, sedimentation, resuspension
- Fate of substances
  - Transport, drifting, spreading, diffusion
  - Process modelling
- Climate change scenarios
  - Multi-year modelling

SYKE

Basic lessons

- Solution orientation ➔
  - Model to be applied has to solve the problem at hand
  - As simple model as possible
  - We should have good understanding what happens in nature
- Model (=Tool) known and at hand
  - Well documented
  - Open source code
  - Possibility to modify the model
  - Collective experience and team work
- Data collection
  - Calibration
  - Validation

SYKE

Related literature

1. Huttula T., 1994: Suspended sediment transport in Lake Säkylän Pyhäjärvi. Aqua Fennica, Vol. 24,2: 171-185
2. Huttula T., Peltonen A., Bilaletdin Ä., Saura M., 1992: The effects of climatic change on lake ice and water temperature. Aqua Fennica Vol. 22,2.
3. Myrberg, K., Ryabchenko, V., Isaev, A., Vankevich, R., Andrejev, O., Bendtsen, J., Erichsen, A., Funkquist L., Inkala A., Neelov I., Rasmus K., Rodriquez Medina M., Raudsepp U., Passenko J., Söderkvist J., Sokolov A., Kuosa H., Anderson T. R., Lehmann A. & Skogen M. D. 2010: Validation of three-dimensional hydrodynamic models of the Gulf of Finland. Boreal Env. Res. vol. 15, nº5, pp. 453-479
4. Huttula T., Pulkkanen M., Arkhipov B., Leppäranta M., Solbakov V., Shirasawa K., and Salonen K., 2010: Circulation in an ice covered lake. Estonian Journal of Earth Sciences, 2010, 59, 4, 298-309.
5. Rankinen, K., Valpasvuo-Jaatinen, P., Karhunen, A., Kenttämies, K., Nenonen, S. and Bärlund, I. 2009. Simulated nitrogen leaching patterns and adaptation to climate change in two Finnish river basins with contrasting land use and climatic conditions. Hydrology Research 40(2-3)177-186.
6. Veijalainen, N., Lotsari E., Alho, P., Vehviläinen, B., Käyhkö, J. 2010. National scale assessment of climate change impacts on flooding in Finland. Journal of Hydrology 391: 333-350.
7. Bärlund, I., Rankinen, K., Järvinen, M., Huitu, E., Veijalainen, N. and Arvola, L. 2009. Three approaches to estimate inorganic nitrogen loading under varying climatic conditions from a headwater catchment in Finland. Hydrology Research 40(2-3): 167-176.

SYKE

SYKE

Future prospects

Hiroshi Suito[1] and Timo Huttula[2]

[1]Okayama University, Japan
[2]Finnish Environment Institute, Finland

As presented in this report, various topics related to aquatic problems have been addressed by researchers of OU and SYKE in this project. During this project period, our collaboration has been related mainly to scientific approaches, but it has not remained restricted to them. We have enjoyed mutual understanding through the scientific discussion and social events that have taken place during our reciprocal visits. We have sought mutual understanding in terms of culture and history, which has deepened our collaborations and has made them much more fruitful.

This project officially ends in June 2014, but we regard that event as a starting point for our next step forward. Now we have come to understand that there are many not only important but also interesting problems that we share. We are poised to proceed to wider and deeper research collaborative efforts.

Furthermore, several young scientists and graduate students have participated in this project. Especially for them, this project has provided an invaluable experience. We believe they will extend international collaborative efforts of this kind throughout their scientific careers.