Klöve B., J. Kettunen, O. Varis, P. Vakkilainen (1993) A real-time monitoring system for Kerava river quality.
Environmental Monitoring and Assessment (24), 167-180.

A REAL-TIME MONITORING SYSTEM FOR KERAVA RIVER QUALITY

B. KLØVE, J. KETTUNEN, O. VARIS, P. VAKKILAINEN and H. SIRVIÖ
Helsinki University of Technology, Laboratory of Hydrology and Water Resources Management, SF–02150 Espoo, Finland

(Received: May 1991)

Abstract. The hygienic quality of the water of the Kerava river, southern Finland, deteriorates occasionally. The purpose of the study was to design a real-time monitoring system that would inform the public using the river for recreational purposes about the changes in water quality. The system was constrained to consist of on-line sensing of water quality and quantity, and adjacent forecasting models. Four different system alternatives were analyzed and compared. The first alternative observes river flow in real-time; the second alternative also monitors water temperature, turbidity, pH, conductivity and dissolved oxygen. The data collected in this way are used to forecast *Streptococcus* and *E. coli* concentrations, using canonical correlation and regression analysis. The third configuration is a two-step procedure, where river flow is first predicted by an ARMAX model and the hygienic state is then based on the flow estimate, as in the first assemblage. The most expensive monitoring system, which at present is the least well-known, is to apply the Lidar system, where the hygienic status of the river quality is observed directly using laser technology, placing less emphasis on modeling. In this paper, the alternatives are formulated and a preliminary comparison is made, using the criteria of operational feasibility, prediction uncertainty, investment and maintenance costs, and suitability for in-situ monitoring.

1. Introduction

Substantial restoration measures to improve the water quality and recreational utility of the Kerava river were undertaken in the 1980s. A sewage transfer tunnel was constructed to remove most of the municipal waste water from the catchment, the river water was diluted with additional discharge from another catchment, and the shores of the river were restored so as better to serve recreational activities. These measures have yielded considerable improvement in the recreational utility of the river. However, the hygienic state of the river water still presents a problem from time to time, and this was the focus of the present study.

The abundance of microbes is only monitored regularly in the summer. A sample is taken weekly, incubated for up to 48 h, and then analyzed. The problem is that information on the hygienic state of the river water is delayed, usually by several days or even weeks. Because the river water quality is subject to rapid changes owing to the flow and quality regimes, such delays substantially impair the utility of the entire monitoring system. The scope of this study was to analyze the possibilities for designing a real-time system for monitoring the river water quality. The approaches chosen include prediction of microbes using models based on the correlation of off-line observational data and inorganic variables measurable on-line, and direct real-time monitoring of microbes using the Lidar system.

Environmental Monitoring and Assessment **24:** 167–180, 1993.

2. Site Description and Data

2.1. KERAVA RIVER

The Kerava river, southern Finland, is a small, eutrophic river with a catchment area of 397 km^2; it is 50 km long and 4–10 km wide. The high percentage of clay soils is the natural reason for the eutrophy of the river. This characteristic is strongly enhanced by non-point pollution from agriculture and non-urban areas.

Because of the small catchment area and the scarcity of reservoirs, the water flow is subject to rapid alterations. The river flow may change during summer months, from the average of 1 m^3/s to as much as 20 m^3/s within one day. The increased flow makes the water turbid and increases the concentrations of nutrients and pathogenic bacteria. In consequence, the water quality deteriorates for recreational purposes, particularly for swimming. The average quality of the river is classified as fair by national standards. A general feature is the gradually increasing deterioration in water quality from upstream to downstream.

2.2. THE DATA FOR THIS STUDY

This study made use of monitoring data collected between 1980 and 1990. The gauge sites are shown in Figure 1. The water quality variables monitored include water temperature, oxygen, turbidity, suspended matter, conductivity, pH, color, COD, BOD_7, total nitrogen, NO_3-N, NO_2-N, NH_4-N, total phosphorus, PO_4-P, and the abundance of two hygienic indicators, *Streptococcus* and *Escherichia coli*. The bacteria counts were determined weekly throughout the swimming season, i.e. from June to August (Figure 2). The other variables were measured more frequently. The water quality data from 1980–1987 were obtained from the Environmental Information Center of the National Board of Waters and the Environment, and from 1987–1990 from local municipalities (Keski-Uusimaa Inter-Municipal Corporation for Water Protection).

One observation, yielding concentrations of 100 000 *Streptococcus* cells/100 ml and 18 000 *E. coli*/100 ml, was deleted as an outlier from the time series of site K35. These values are very exceptional and may indicate a leak in the sewerage system, and is thus not an output of the system being modeled.

Attention has recently shifted from monitoring nutrient levels to monitoring the abundance of hygienic indicator bacteria. From 1985–1988, research activities concentrated on investigating the relationship between the monitored indicator bacteria and the actual pathogenic bacteria, and the variation in the abundance of bacteria. Niemi *et al.* (1987) detected a linear relationship between *E. coli* and the pathogenic bacteria, but the relation between *Streptococcus* and pathogens was not so straightforward. Niemi and Niemi (1988) discuss the relationship between river flow and the amount of bacteria.

Precipitation data from 1977–1981 collected by the Institute of Meteorology were also used. The areal precipitation was calculated as the average for four gauge sites. The river flow data were obtained from the Hydrological Office, National

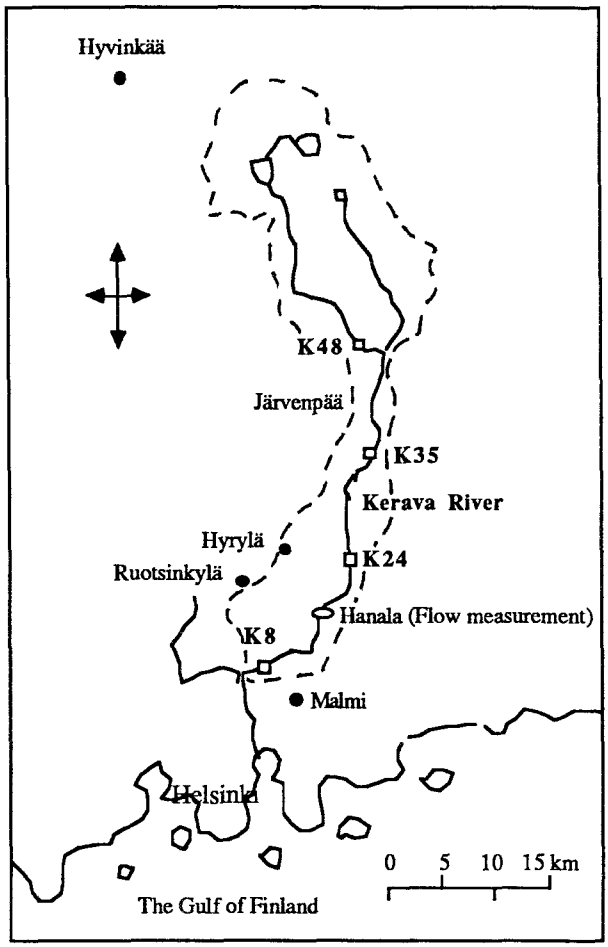

Fig. 1. Location of gauge sites for measuring water quality (K8, K24, K35 and K48), river flow (Hanala), and precipitation (black dots). The number after the K represents the distance, in kilometers, from the discharge point of the river to the Baltic Sea. Broken line shows the drainage basin outline.

Board for Waters and the Environment. The flow was measured at Hanala, which is located between gauge sites K8 and K24. The hydrometeorological variables, river flow and precipitation, were measured daily, the flow as the daily average and the precipitation as the daily sum of rainfall.

3. On-Line Monitoring of Water Quality

3.1. Direct and Indirect Monitoring of Bacteria

Real-time forecasts on bacteria can be based on either direct monitoring of bacteria or sensoring of automatically measurable water quality variables, and followed by

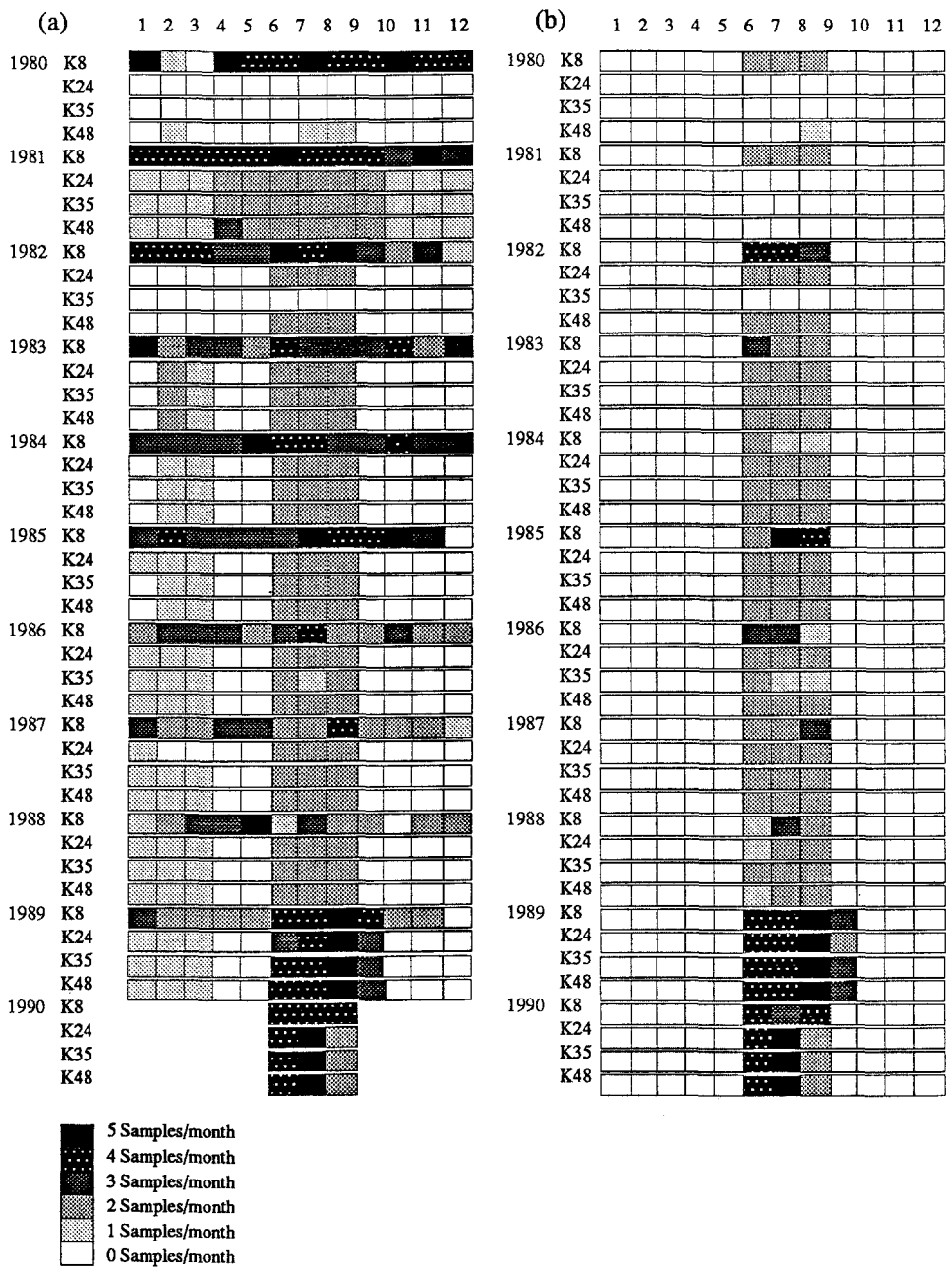

Fig. 2. Monthly monitoring frequency at the four gauge sites between 1980 and 1990. (a) Water quality monitoring; (b) monitoring of hygienic indicators.

forecasting of the abundance of bacteria using an empirical relationship between bacteria and water quality variables. An on-line measuring system consists of a measuring device and a system for transmitting and receiving data. The digital signal obtained from the sensors can be transmitted by either radio or telephone lines.

3.2. REAL-TIME MONITORING FACILITIES

Two different approaches to in-situ water quality monitoring are discussed, one using transducers and the other the Lidar technique. The photometric measurement of nutrients is not discussed here because the on-line analytical instruments available are not suitable for analyzing river water quality.

Most commonly, on-line measurements of the inorganic water quality of natural waters are obtained by using transducers. The transducer operates on the principle that the substance being measured induces a physical or chemical change. The change is converted into an electric voltage or current, the magnitude of which is related to the substance being measured. The potentiometric measurement of inorganic substances has many important advantages (Hulanicki and Trojanowicz 1979). The relationship between electric potential and concentration – or activity, to be more precise – is basically logarithmic. This makes it possible to measure concentrations over several orders of magnitude with constant precision. The response time is short, varying from seconds to minutes, thus facilitating *in-situ* river quality monitoring when the concentrations are subject to rapid changes.

When used for river water quality monitoring purposes, the transducer technique can analyze water flow, precipitation, pH, oxygen, conductivity, redox and chloride. NH_4-N and NO_2-N can also be measured, but the precision is not good enough for river water quality monitoring. Nitrogen probes have been discussed in more detail by Virtanen (1981). Kløve (1991) discussed the performance and prices of various water quality measuring transducers. Grabner *et al.* (1986) investigated the performance of a phosphate-sensing probe. The measurement of bacteria by a bioelectrode is under investigation (Corcoran and Kobos 1987).

The Lidar system is the only equipment included in this study which is applicable to the real-time monitoring of organic compounds. The system is based on measurement of the fluorescence emitted by the organic compounds after they have been exposed to a laser beam. By varying the amplitude of the laser light according to the organic substance to be measured, a specific emission can be obtained, which is compared with the substance's emission image computerized beforehand. Each concentration has its own image, which has to be recorded in the computer memory. The recording requires a number of chemical analyses of the water to be made.

TABLE I
The four system assemblages formulated and investigated.

Method	Input variables	application
1a	Flow	Indirect forecasting of the amount of bacteria
1b	pH, oxygen, conductivity, temperature, turbidity, and flows	Indirect forecasting of the amount of bacteria
2	Flows and precipitation	Indirect forecasting of the amount of bacteria
3	–	Direct monitoring of the amount of bacteria

4. Predictive Models and Monitoring System Assemblages

There is evidently a variety of possibilities for forecasting and monitoring the water quality of the Kerava river. In practice, however, the present system was restricted to on-line water quality sensoring with adjacent forecasting models or the Lidar technique. Four different alternative assemblages (Table I) summarizing the findings described below in more detail were created.

The most appropriate indirect forecasting approach was chosen using the two-step procedure shown in Figure 3. The criteria on which the water quality variables were selected for the indirect forecasting of bacteria were that the variables had to be correlated with bacteria, and equipment with sufficient precision and suitability for real-time monitoring had to be available. The relationship between bacteria and the water quality variables chosen was then studied in more detail. The aim of the procedure was to find the combination of water quality variables best able to explain the variation in the abundance of bacteria.

4.1. Correlations between water quality variables

The correlations of water quality variables with the bacteria at four different gauge sites are shown in Figure 4. Turbidity had the strongest correlation; the next strongest was river flow. The other variables showed a weaker correlation with the bacteria. The correlation varied spatially, being strongest at sites K24 and K48 and weakest at site K35, where nearly all of the variables showed some correlation with *E. coli* and *Streptococcus*. The water quality variables were mutually correlated (Figure 5).

4.2. Canonical correlation analysis for diagnosis and prediction

The correlation structure was studied further with canonical correlation analysis. The procedure used was that of Varis (1991), in which the canonical correlation between two sets of normalized and standardized data is first calculated, and a linear model between canonical variates and the variables in the set to be predicted

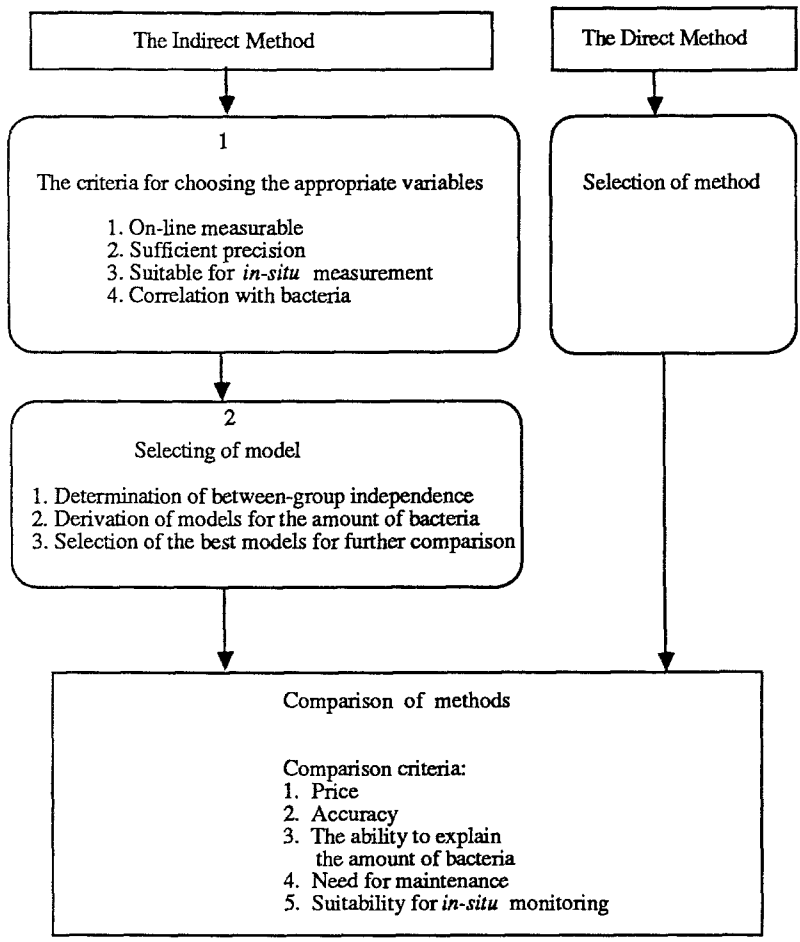

Fig. 3. Procedure for determining the alternatives for the river monitoring system.

is then estimated. Since the data have a mean of zero and are of unit variance, the correlations between canonical variates and original variables, usually presented as diagnostic plots (Figure 6), equal the regression coefficients of the high-resolution predictive model. These diagnostic plots can be used to illustrate the correlation structure of the original variables, in terms of both strength and direction. When interpreting the correlation strengths, one must also take the statistical significance of the respected canonical variates into account. The canonical variates are uncorrelated, owing to the principle of canonical correlation analysis (e.g. Giri 1977, Gnanadesikan 1977, Gittins 1985), and thus an unbiased multivariate regression model can be identified straightforwardly.

The diagnostic results of the canonical correlation analysis (Figure 6) stress the

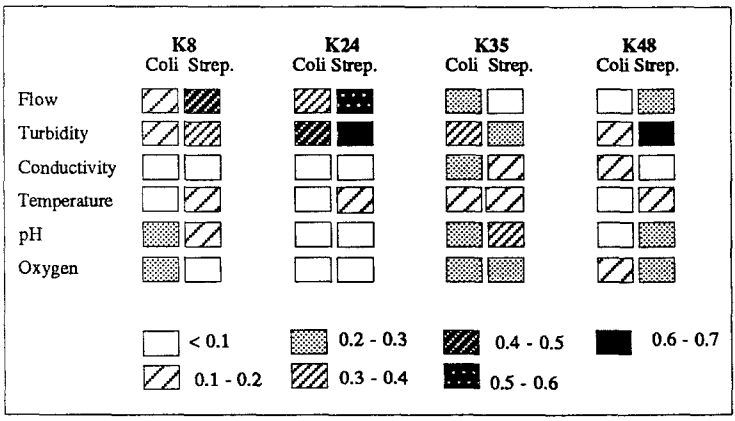

Fig. 4. Correlation strengths (Pearson's r) between water quality variables including water flow at site K8.

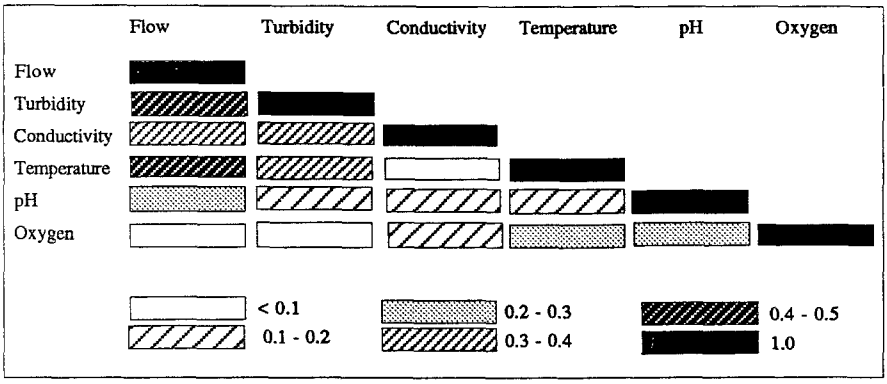

Fig. 5. Correlation strengths (Pearson's r) between water quality variables including water flow at site K8.

importance of river flows in forecasting. Consequently, two types of model were made at each site, one (1a) using the canonical variates made from the river flow and the other (1b) using the canonical variates made from all the possible variables. Figure 6 shows the results of the latter analysis. At all sites, river flow, turbidity, and *Streptococcus* correlated in the same direction. *E. coli* and *Streptococcus* correlated roughly in the same direction; thus both bacteria indeed indicate the same situation. Oxygen and water temperature correlated at all sites in the opposite direction to the amount of bacteria. The first canonical variate explained about 70% of the variation in bacteria (Tables II and III).

The predictive regression models for *E. coli and Streptococcus* were:

$$x_i = a_i v_1 + b_i v_2 \tag{1}$$

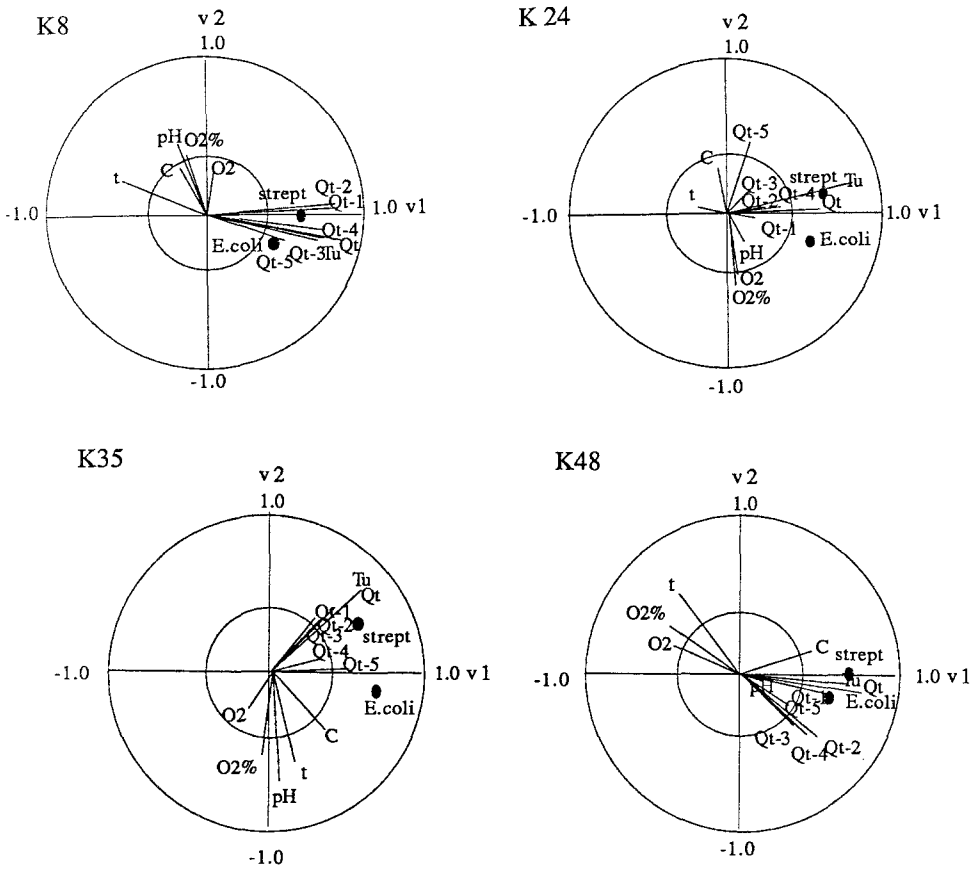

Fig. 6. Diagnostic plots of the canonical analysis for system assemblage 1b: Correlation between original variables and canonical variates. The inner origo centered circles indicate the 99.9% confidence level. Q denotes flow, with time delays up to 5 days, t stands for temperature, Tu for turbidity, C for conductivity, O2 for dissolved oxygen concentration, O2% for dissolved oxygen percentage of saturation, and strept for *Streptococcus*.

TABLE II

Canonical correlation coefficients (Pearson's r) at the four gauge sites with system assemblage 1a. $CC1$ stands for the first and $CC2$ for the second canonical correlation. $P(CC1 = 0)$ indicates the probability that the first canonical correlation is zero.

Gauge site	The canonical correlation coefficients			
	$CC1$	$P(CC1 = 0)$	$CC2$	$P(CC2 = 0)$
K8	0.48	0.006	0.31	0.17
K24	0.50	0.019	0.31	0.29
K35	0.61	0.001	0.39	0.12
K48	0.63	0.07	0.21	0.82

TABLE III
Canonical correlation coefficients at the four gauge sites with system assemblage 1b. $CC1$ stands for the first and $CC2$ for the second canonical correlation.

Gauge site	The canonical correlation coefficients			
	$CC1$	$P(CC1 = 0)$	$CC2$	$P(CC2 = 0)$
K8	0.58	0.0082	0.40	0.28
K24	0.70	0.003	0.48	0.18
K35	0.75	0.0001	0.53	0.13
K48	0.70	0.008	0.37	0.82

TABLE IV
Parameters of the predictive regression models at the four gauge sites with system assemblage 1a.

Site	Observations (No.)	Bacteria	a	b	R^2
K8	84	*E. coli*	0.369	-0.191	0.18
		Streptoc.	0.458	0.087	0.22
K24	67	*E. coli*	0.425	-0.166	0.21
		Streptoc.	0.470	0.112	0.23
K35	59	*E. coli*	0.433	-0.256	0.27
		Streptoc.	0.585	0.082	0.36
K48	55	*E. coli*	0.519	0.119	0.28
		Streptoc.	0.580	-0.083	0.34

where v_1 and v_2 are the first and second variates, respectively, a_i and b_i are regression parameters, and x_i denotes the normalized and standardized observation of the bacteria group i, in this case either *E. coli* or *Streptococcus*.

Some statistics of the models of types 1a and 1b are listed in Tables IV and V. The performance of the model was tested by forecasting the values on which the model had been based and comparing them with the actual values.

4.3. USING FLOW FORECASTS AS THE BASIS FOR CANONICAL MODELS

The third set of predictive models (2) was based on time-domain flow forecasts. This was done first by forecasting the river flow with an ARMAX model and using the resulting forecast as an input in models similar to those in 1a. The output is the forecasted amount of bacteria. The ARMAX model was obtained using data for the summer of 1981, which was a rainy year. The performance of the model was tested by comparing the one-day forecast and the measured flows.

TABLE V
Parameters of the predictive regression models at the four gauge sites with system assemblage 1b.

Site	Observations (No.)	Bacteria	a	b	R^2
K8	84	*E. coli*	0.389	-0.293	0.24
		Streptoc.	0.571	0.06	0.33
K24	67	*E. coli*	0.580	-0.270	0.41
		Streptoc.	0.667	0.154	0.47
K35	59	*E. coli*	0.695	0.164	0.54
		Streptoc.	0.598	0.310	0.46
K48	55	*E. coli*	0.545	-0.249	0.36
		Streptoc.	0.717	0.098	0.52

4.4. Comparison of the Monitoring System Alternatives

The results of a preliminary comparison of the four alternative monitoring system assemblages (Table I) are presented in Table VI, which summarizes the properties of these four options with respect to five criteria (Figure 3) chosen. Price here represents the purchase price of the analyzing equipment. Accuracy means the relative measurement accuracy and its effect in the model. The analysis errors were assumed to be mutually independent. The relative error of the analysis was used as an estimate for the variance. The definition of maintenance is obvious. Suitability is defined as the suitability of the analysis for *in-situ* monitoring of river quality compared with laboratory measurements. To achieve reliable interpretation in the comparison, the different criteria were assumed to be independent. This multicriteria decision problem involving a comparison of these four alternatives in terms of costs and forecast uncertainties within the framework of risk attitude analysis has been discussed in more detail by Varis *et al.* (1992).

5. Discussion and Conclusions

Fluorescence is widely used in the analysis of microbes in medical research, e.g., in fluorescence microscopy. Fluorescence spectrophotometry is also commonly used in water quality studies to analyze organic compounds. Further case-specific studies are needed to determine the sensitivity of the Lidar system in inducing a microbe-specific fluorescence which can be detected from the fluorescence of the background compounds of the water of the Kerava river. If this were possible, the Lidar system would most definitely be a feasible approach with very high potential, especially in comparison wiht the transducer approach, which is more labor-intensive in terms of service and maintenance.

The utility of hydrologic observations, i.e. flow and precipitation data, appeared

TABLE VI
Comparison between the four system assemblages with respect to five selected criteria.

Method	Comparison criteria				
	Price (FIM)	Accuracy	Forecast	Maintenance	suitability for *in-situ* measurements
1a	3 500	* * *	* * *	* * * *	* * * *
1b	55 000	* *	* * * *	* *	* *
2	10 000	*	* *	* * *	* * *
3	700 000	* * * *	* * * * *	* * * * *	*

Accuracy	* * * * *	:≈ < 5%	Forecast	* * * * *	:≈ Very good
	* * * *	:≈ 5 – 10%		* * * *	:≈ Good
	* * *	:≈ 10 – 15%		* * *	:≈ Fairly good
	* *	:≈ 15 – 20%		* *	:≈ Fair
	*	:≈ 20 – 25%		*	:≈ Poor
Maintenance	* * * * *	:≈ None	Suitability	* * * * *	:≈ Very good
	* * * *	:≈ Automatic	for *in-situ*	* * * *	:≈ Good
	* * *	:≈ 1/month	measurements	* * *	:≈ Fairly good
	* *	:≈ 1/week		* *	:≈ Fair
	*	:≈ 1-2/week		*	:≈ Poor

to be high, perhaps higher than was expected before the study. On-line monitoring of these variables is inexpensive and would enable monitoring more often than the present daily frequency.

Instead of modeling *E. coli* and *Streptococcus* separetely, their first canonical variate which is their linear sum, could also be modeled as a function of v_1. This approach would be reasonable especially for practical purposes, when the primary interest is not the individual abundances of *E. coli* and *Streptococcus*.

The models are based on the correlation structure existing between the bacteria and the water quality variables. Turbidity, and not water flow, had the strongest correlation with the bacteria. This is in apparent contradiction to the fact that turbidity is caused by river flow. The explanation for this is straightforward: Turbidity has been measured at the same time and spot as the bacteria, whereas the river flows are measured at just one site and not at the same time as the bacteria.

Interpretation of the negative correlation between the variables indicating productivity – pH, water temperature and oxygen – and the bacteria concentration and river flow is not straightforward. The pathogenic bacteria originate from human and animal residue, and thus are not capable of reproduction in an environment

such as river water, where their numbers are always doomed to fall. Apparently during high biological production in the river, the rate of falling is enhanced. The negative correlation between the abundance of bacteria and the variables indicating productivity could be due to the fact that the abundance of bacteria is positively correlated with low river flow rates. When the flow rate is low the dilution is small, increasing the possibility of a rise in the variables indicating productivity.

The monitoring system alternatives were compared on the basis of knowledge derived from this study. No attention was paid to improving the assemblage or to reducing the uncertainty involved. The measurement inaccuracy of assemblage 2, for instance, is mostly due to the inaccuracy of the precipitation measurement. This inaccuracy could be decreased by increasing the number of precipitation measurement sites.

In the present study, the real-time monitoring in a river used mainly to recreational purposes is discussed. Four different system assemblages were analyzed computationally, and their preliminary comparison was performed using five criteria affecting the practical feasibility of the system. A more detailed analysis of this multicriteria problem are documented and recommendations are presented by Varis *et al.* (1992), where probabilistic modeling, risk analysis, and optimization are being used.

Acknowledgements

This study was funded by the Keski-Uusimaa Inter-Municipal Corporation for Water Protection, by the Academy of Finland, and by the MAVERO project. Special thanks are due to Mr. I. Hirsto and Mr. M. Pekkarinen, and to all our other colleagues who helped us to carry out this study.

References

Corcoran, C.A. and Kobos, R.K.: 1987, 'Selectivity Enhancement of an *Escherichia coli* Bacterial Electrode Using Enzyme and Transport Inhibitors', *Biotechnology and Bioengineering* **4**, 565–570.

Giri, N.C.: 1977, *Multivariate Statistical Inference*, Academic Press, New York.

Gittins, R.: 1985, *Canonical Analysis – A Review with Applications in Ecology*, Springer-Verlag, Berlin.

Gnanadesikan, R.: 1977, *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, New York.

Grabner, E.W., Vermes, I. and König, K.-H.: 1986, 'A Phosphate Sensitive Electrode Based on $BiPO_4$-Modified Glassy Carbon', *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry* **214**, 135–140.

Hulanicki, A. and Trojanowicz, M.: 1979, 'Application of Ion-Selective Electrodes in Water Analysis', *Ion-Selective Electrode Review* **1**, 207–250.

Kløve, B.: 1991, 'A Real-Time Forecasting and Measuring System for the Quality of River Kerava', M.Sc. Thesis, Laboratory of Hydrology and Water Resources Management, Helsinki University of Technology, Espoo.

Niemi, J. and Niemi, M.: 1988, 'The Hygiene Project', *Vantaa River Yearbook 1988* (in Finnish), Keski-Uusimaa Inter-Municipal Corporation for Water Protection, Vantaa, pp. 8–11.

Niemi, M., Niemi, J. and Aarnio, T: 1987, 'Indicators of the Hygienic State in Vantaa and Kerava Rivers' (in Finnish), National Board of Waters and the Environment, Mimeograph Series 37, Helsinki.

Varis, O.: 1991, 'A Canonical Approach to Diagnostic and Predictive Modelling of Phytoplankton Communities', *Archiv für Hydrobiologie* **122**, 147–166.

Varis, O., Kløve, B., and Kettunen, J.: 1992, 'Evaluation of a Real-Time Forecasting System for River Quality – A Trade-off Between Risk Attitudes, Costs and Uncertainty', in press.

Virtanen, R.: 1981, 'Evaluation of Commercial Ion Selective Electrodes', Technical Research Center of Finland, Publications 6/1981, Espoo.