

A Simple Algorithm for On-Line Prediction of BOD₅ by a Microprocessor-Based System

Roni Raviv and Sam Ben-Yaakov*

Department of Electrical and Computer Engineering,
Ben Gurion University of the Negev, Beer Sheva, Israel

Accepted for Publication July 12, 1983

A novel algorithm for predicting BOD₅ from the dissolved oxygen (DO) concentration after a relatively short incubation period is presented and evaluated experimentally. Test runs on synthetic and experimentally derived raw data suggest that BOD₅ can be predicted to within 15% in ca. 36 h. The method can be improved by filtering out, via a digital filter, noise from the raw data. The suggested algorithm does not require elaborate computations or large data storage and can therefore be implemented on a low-cost microcomputer for fast on-line determination of BOD₅.

INTRODUCTION

Despite the open controversy regarding its precision and reliability,¹ BOD₅ is still considered one of the fundamental water quality parameters.² As such, it is widely used to assess organic loading of sewage and treated waters. The standard method for BOD₅ determination³ is cumbersome and time consuming and introduces at least a five-day delay between the instant of sampling until the BOD₅ value is obtained. A number of investigators have attempted to reduce the long delay by suggesting modified analytical techniques to determine BOD₅ indirectly.^{4,5}

One of the early attempts to determine BOD₅ indirectly by an analytical method that lasts only a few hours is based on a precalibration procedure which correlates BOD₅ with dissolved oxygen (DO) consumption during the first few hours of BOD incubation.⁵ However, since a calibration graph must be prepared for each type of water used, the method is useless when sample to sample variability is expected. Furthermore, since the method is based only on the first phase of DO consumption, it is in fact attempting to correlate BOD₅ with the concentration of most easily digestible fraction of the organic load. Hence, any change in the percentage of this fraction from the total, as well as a change in the distribution of the microorganisms in the original sample, will invalidate the calibration graph.

A second method for determining BOD₅ indirectly is based on chemical oxygen demand (COD) analysis and a predetermined calibration curve which correlates BOD with COD.⁴ It has been documented, however, that the correlation of BOD and COD is not high and is specific to each type of water tested.^{1,6} Recent developments in microelectronic technology and, in particular, the introduction of low-cost microcomputers have made possible the practical realization of "smart" analyzers that can not only supervise the instrumentation system but also perform rather sophisticated data processing on line.⁷

The purpose of the present study is to investigate the possibility of using such a system for the determination of BOD₅ in a shorter time than the standard method, which lasts five days. This can be accomplished if the DO consumption during the BOD₅ incubation period is approximated by a mathematical model which can then be used to predict the DO concentration after five days (i.e. BOD₅) from the initial stage of the process. Such a method will be practical if 1) the DO consumption during the BOD incubation period could indeed be described consistently by a mathematical model, and 2) if the suggested prediction algorithm is simple enough (i.e. does not require elaborate computation and large memory) to be implemented on a microcomputer.

MODEL AND ALGORITHM

Numerous investigators have concluded that the organic material utilization process by heterotrophic microorganisms can be approximated to a high degree of correlation by a first-order kinetics⁸⁻¹¹:

$$dS(t)/dt = -K(S)S(t) \quad (1)$$

where $S(t)$ is the organic material present at time t in terms of potential oxygen demand (mg/L); $K(S)$ is the deoxygenation constant which may vary during the process (h^{-1}); and t is the time (h).

Equation (1) assumes endogenous respiration,¹ although growth may take place during the first few hours of the BOD incubation period.

*To whom all correspondence should be addressed.

The total BOD (BOD_L) or ultimate oxygen demand (UOD) is the total available organic materials (in terms of potential oxygen consumption) available for the micro-organism. Hence,

$$S(0) = BOD_L \triangleq DO(0) - DO(\infty) \quad (2)$$

where $DO(0)$ is the initial dissolved oxygen concentration in sample, and $DO(\infty)$ is the final DO in the sample after all the digestible organic matter has been consumed. Consequently, the amount of digestible organic matter presented at time t , $S(t)$ is

$$S(t) = S(0) - [DO(0) - D(t)] \quad (3)$$

The total digestible organic matter can now be related to the DO at any given time t , by combining eqs. (1) and (3):

$$S(0) = -\frac{1}{K(S)} \frac{dDO(t)}{dt} + DO(0) - DO(t) \quad (4)$$

which can be rearranged in the form:

$$\frac{dDO(t)}{dt} = -K(S)DO(t) - K(S)[S(0) - DO(0)] \quad (5)$$

Assuming that $K(S)$ is constant over the range of interest, and using least squares fitting to a straight line of the form

$$y = ax + b \quad (6)$$

where

$$\begin{aligned} a &= -K(S) \\ b &= -K(S)[S(0) - DO(0)] \end{aligned}$$

a and b can be evaluated, from which estimated value of K and $S(0)$ can be computed:

$$\hat{K} = -\hat{a} \quad (7)$$

$$\hat{S}(0) = (\hat{b}/\hat{a}) + DO(0)$$

Inserting the estimated values into the solution of eq. (1) yields

$$\hat{S}(t) = \hat{S}(0) \exp(-\hat{K}t) \quad (8)$$

At the end of the BOD₅ incubation period (120 h), we have

$$\hat{S}(120) = \hat{S}(0) \exp(-120\hat{K}) \quad (9)$$

Combining eqs. (3), (8), and (9), we obtain the prediction formula for \hat{BOD}_5 at time t :

$$\hat{BOD}_5 = DO(0) - DO(t) + S(\hat{0}) [e^{-\hat{K}t} - e^{-120\hat{K}}] \quad (10)$$

Alternately, \hat{BOD}_5 can be expressed as:

$$\hat{BOD}_5 = \hat{S}(0) [1 - e^{-120\hat{K}}] \quad (11)$$

The \hat{BOD}_5 prediction by eq. (10) should yield better results than the prediction by eq. (11). This stems from the fact that eq. (10) utilizes the data on the total amount of oxygen consumed up to the prediction time, i.e.

$[DO(0) - DO(t)]$ and predicts only the residual DO consumption to $t = 120$ h, $\hat{S}(0) [e^{-\hat{K}t} - e^{-120\hat{K}}]$. On the other hand, eq. (11) attempts to predict \hat{BOD}_5 from the estimated \hat{K} and $\hat{S}(0)$ without this available information. Since the parameter estimates are not perfect and the model [eq. (1)] is only a first-order presentation of the process:

$$\hat{S}(0)e^{-\hat{K}t} \neq DO(0) - DO(t) \quad (12)$$

the accuracy of BOD₅ prediction by eq. (11) should be poorer. Therefore, eq. (10) was used in this work as the basic relationship for \hat{BOD}_5 prediction.

Examination of estimation equations (4), (6), and (7) and prediction equation (10) reveal that the only data required for predicting BOD₅ by the above suggested procedure is $DO(t)$. Furthermore, estimation of a and b by eq. (6) and hence \hat{K} and $\hat{S}(0)$ [eq. (7)] does not require storage of all data points, since the least-squares estimation can be done recursively, processing the $DO(t)$ data as soon as measured. This feature is especially useful when application of microcomputers is envisaged as the memory requirements are minimal.

Special attention must be paid, however, to the problem of noise which will always corrupt the measured value of $DO(t)$:

$$D\hat{O}(t) = DO(t) + n(t) \quad (13)$$

where $D\hat{O}(t)$ is the measured value of $DO(t)$; $DO(t)$ is the true value of dissolved oxygen at time t ; and $n(t)$ is the noise at time t . Taking the noise into consideration, estimation equation (5) can be rewritten in terms of the measured value $D\hat{O}(t)$:

$$\begin{aligned} \frac{dD\hat{O}(t)}{dt} - \frac{dn(t)}{dt} &= -K [D\hat{O}(t) + n(t)] \\ &\quad - K[S(0) - DO(0) + n(0)] \end{aligned} \quad (14)$$

Since the noise frequency is higher than the signal frequency, one can assume:

$$\frac{dn(t)/dt}{dDO(t)/dt} \gg \frac{n(t)}{DO(t)} \quad (15)$$

Thus, eq. (14) can be approximated by

$$\frac{D\hat{O}(t)}{dt} - \frac{dn(t)}{dt} = -KD\hat{O}(t) - K[SO(0) - D\hat{O}(0)] \quad (16)$$

which now fits the model

$$y = ax + b \quad (17)$$

in which the uncertainty is in y only (Bevington 1969).

It is obvious from the above discussion that the prediction of \hat{BOD}_5 by the suggested method (and in fact by any extrapolation method) could be improved by filtering out the $dDO(t)/dt$ noise. This was accomplished in the present study by prefiltering the $DO(t)$ raw data and by

applying a numerical differentiation method with an inherent smoothing effect. The data was filtered by a second order Butterworth type digital filter with a cutoff frequency of 0.2 L/h which is 0.05 the sampling rate used (4 L/h). Using a standard numerical technique,¹² the filter equation for the i th sample was evaluated as:

$$\begin{aligned} \hat{D}\hat{O}(i) = & 0.0192\hat{D}\hat{O}^*(i) + 0.0385\hat{D}\hat{O}^*(i-1) \\ & + 0.0192\hat{D}\hat{O}^*(i-2) + 1.5712\hat{D}\hat{O}^*(i-1) \\ & - 0.648\hat{D}\hat{O}^*(i-2) \end{aligned} \quad (18)$$

Further smoothing was obtained by using a modified method for differentiation in which the original curve is approximated by a second-order polynomial using least-squares fitting.¹³ The equation of the i th value for a backward numerical differentiation was evaluated to be:

$$\begin{aligned} \frac{d\hat{D}\hat{O}(i)}{dt} = & [1/20(0.25)] [21\hat{D}\hat{O}(i) - 13\hat{D}\hat{O}(i-1) \\ & - 17\hat{D}\hat{O}(i-2) + 9\hat{D}\hat{O}(i-3)] \end{aligned} \quad (19)$$

EXPERIMENTAL

Experimental Assembly

Experiments were carried out in 1-L dark glass bottles placed in a thermostated bath which was regulated by a dip-type thermostat to ca. $\pm 0.1^\circ\text{C}$ (Fig. 1). A DO sensor¹⁴ was used to monitor the test solution during the BOD incubation period. The DO data was collected by a data acquisition system¹⁵ which recorded the information on a magnetic cassette at a sampling rate of four readings per hour.

Reagents and Solutions

All reagents were analytical grade and prepared according to the standard method of BOD analysis.³ Samples were collected from local sewage treatment plants (Hazerim and Beer Sheva, Israel), filtered, diluted a 1000-fold, and aerated to saturation.

Analytical Procedure

The analytical procedure followed that of the standard method of BOD analysis,³ except for the temperature of incubation which was 30°C instead of the standard 20°C . This deviation was deemed necessary to allow for effective thermostating by heating alone as a refrigeration-heater type thermostat was not available at the time.

Computation

The on-line estimation and prediction algorithm was simulated on a CBM 3032 (Commodore) microcomputer programmed in BASIC to execute the data processing computation. The raw data were first read from the magnetic cassette into diskette using the microcomputer and

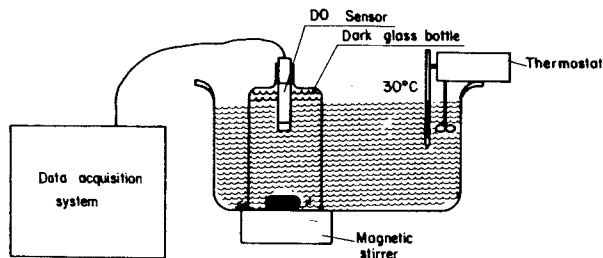


Figure 1. Experimental assembly for DO measurement during BOD incubation.

a dual diskette driver type CBM 3040 (Commodore) and an in-house-built cassette reader.¹⁵

The BASIC estimation-prediction program used the diskette as the raw data source. It should be emphasized that the BASIC program does not require the storage of all the data points and uses the current DO values and updated variables.¹⁶ Hence, with a proper interface (e.g. similar to the one described in ref. 7), the same program would perform the estimation/prediction computation on-line.

The average time required to perform estimation/prediction computation on a single data point was 3 s.

RESULTS AND DISCUSSION

A test run was first performed on synthetic data which was derived from the basic process equation, eq. (8), using the following numerical relationship:

$$\hat{D}\hat{O}^*(i) = 6.00 \exp(-i/240) + 1.00 \quad (20)$$

i.e., an $S(0)$ value of 6.00 [mg/L] and $K = 4/240 = 1/60$ [h^{-1}]. The estimation/prediction procedure was commenced after 15 h of assumed incubation to allow for a stabilization time that was found to be important when processing real data as discussed later.

Results of the test run of synthetic data (Fig. 2) demonstrate the effectiveness of the suggested algorithm in estimating the process parameters $\hat{S}(0)$ and \hat{K} and predicting the \hat{BOD}_5 value. Only few points of the synthetic data are required to obtain a good prediction, and additional points do not improve the predicted value in this ideal case.

An insight into the possible effect of nonideal behavior was obtained by adding white noise to the theoretical process curve. Addition of pseudorandom white noise with a standard deviation of 0.02 mg/L [equivalent to 0.3% $\hat{D}\hat{O}(0)$] manifested itself as expected by a large derivative noise $d\hat{D}\hat{O}(t)/dt$ (Fig. 3). In this case, the convergence of the estimated parameters, and therefore the predicted \hat{BOD}_5 value, was much slower and required about 20 h data collection before a reasonable prediction (to within 10%) of \hat{BOD}_5 was obtained.

The introduction of the low-pass digital filter [eq. (18)] improved the situation markedly by cutting out high frequency components of $d\hat{D}\hat{O}(t)/dt$ and permitting a much

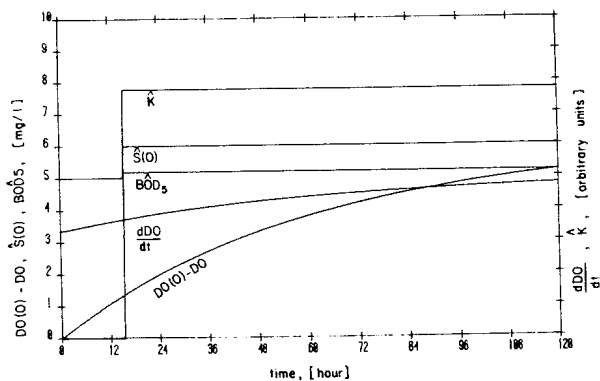


Figure 2. Synthetic data (DO consumed), calculated dDO/dt , estimated \hat{K} and $\hat{S}(0)$, and predicted \hat{BOD}_5 using the suggested algorithm. See text for parameter definitions.

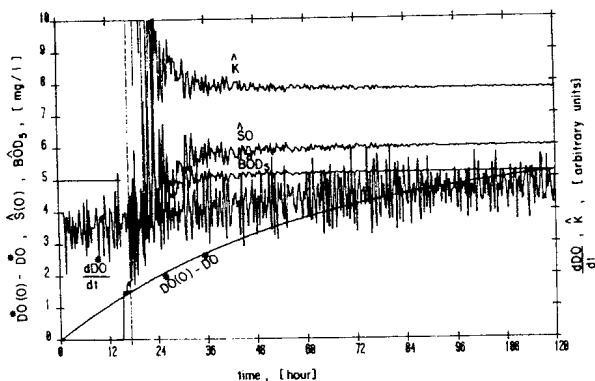


Figure 3. Effect of added white noise of 0.02 mg/L [ca. 0.3% of $DO(0)$], standard deviation to synthetic data on calculated dDO/dt , estimated \hat{K} and $\hat{S}(0)$, and predicted \hat{BOD}_5 .

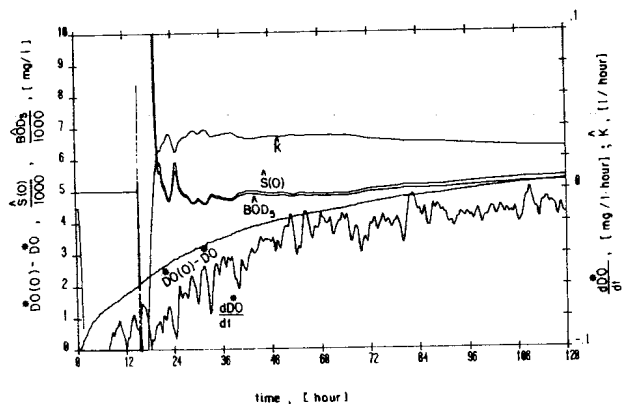


Figure 4. The effect of a low-pass Butterworth-type digital filter on noise-corrupted data of Figure 3 and corresponding signal processing results. The filter's cutoff frequency was 0.2 h^{-1} .

faster convergence. This same behavior was found when a much higher noise level was added and the same filter was used. A white noise level of 0.2 mg/L [ca. 3% of $DO(0)$] filtered by the same digital filter was highly attenuated (Fig. 5), enabling one to obtain a reasonable prediction (to ca. 10%) from a 20-h data segment. This is in a reasonable agreement with the theoretical evaluation

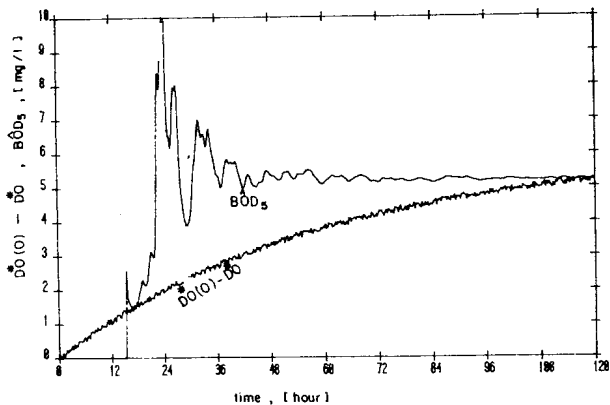


Figure 5. BOD_5 prediction on filtered synthetic DO data corrupted by white noise with a standard deviation of 0.2 mg/L [ca. 3% of $DO(0)$].

of the 90% confidence limits as calculated from the uncertainties in a and b [eq. (6)]. The frequency spectrum of experimentally derived $DO(t)$ data during BOD incubation reveal that most of the BOD information is concentrated at the lower end of the spectrum (Fig. 6). This was used as a justification for locating the cutoff frequency of the digital low-pass filter at 0.2 h^{-1} .

The experimentally derived data also reveals that the first few hours of BOD incubation can not be used for BOD_5 prediction because one clearly recognizes a number of phases until the DO consumption becomes regular (Fig. 7). This behavior was found in all the experimental runs (ca. 12) of this study. It was concluded, therefore, to start the estimation-prediction computation after 15 h delay period since incubation commenced. Using the estimation-prediction algorithm on the experimental data after smoothing it by the digital filter program, a reasonable predicted value of BOD_5 is obtained after processing a data segment of 10 h or ca. 25 h after incubation was commenced. This implies that a reasonable (within say 10%) of BOD_5 can be obtained after only 24 h. This is based on the assumption that the DO consumption process in the test solution stabilizes after 15 h, after which no phase changes are detectable. This validates our assumption that K can be assumed to be constant over the range of interest.

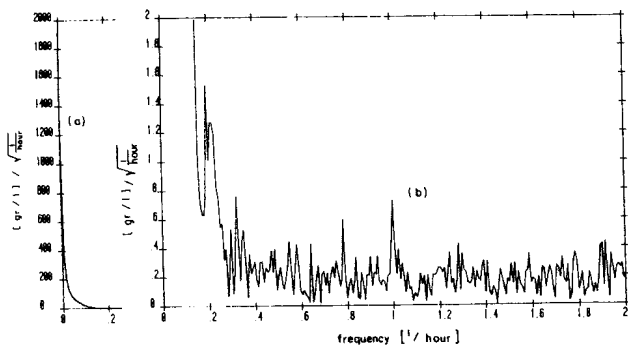


Figure 6. Power spectrum density of typical experimental data (run Z104) shown at (a) low resolution and (b) high sensitivity power scales.

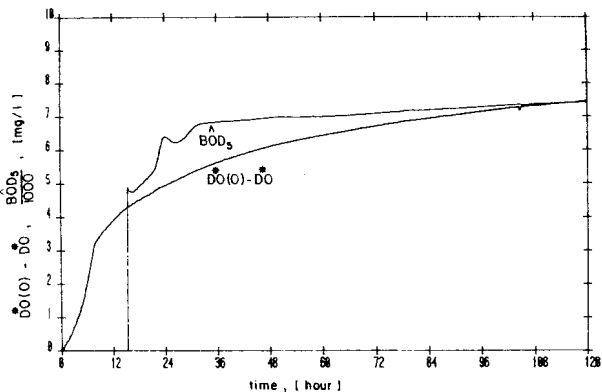
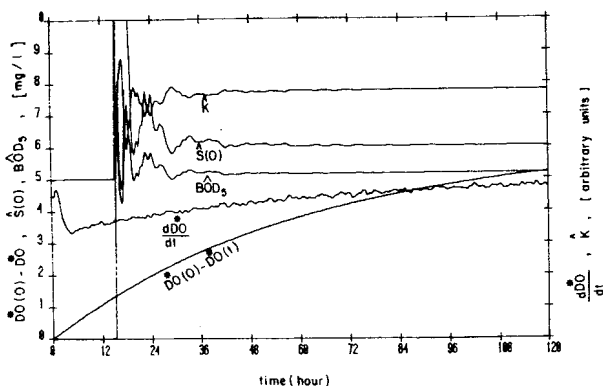


Figure 7. Results of parameter estimation and BOD_5 prediction on experimental data (Z104).

Figure 8. Results of BOD_5 prediction on experimental data (run Z105).

A second example of a typical prediction curve obtained during the present study (Fig. 8) demonstrates again the uncertainty in the initial period of incubation. In this case, the $DO(t)$ consumption curve resembles a growth rate curve and stabilizes to a first-order-like kinetics after ca. 10 h. Acceptable BOD_5 values are also obtained here after ca. 24 h.

The validity of the basic underlying assumption of the present model—that the process can be described by first-order kinetics—can be tested by examining the relationship between dDO/dt and DO . This relationship should be linear if the model can be described by the first-order kinetics equation, eq. (1). As already discussed, the first hours of the processes include several phases and certainly can not be described by the simple model of eq. (1) with a constant reaction rate constant (K). Examination of experimentally derived data for incubation times longer than ca. 15 h reveal (Fig. 9) that a large segment of the process conforms to first-order kinetics. The data presented in Figure 9 suggest that the incubation period between 15 and 120 h (DO levels of ca. 5.2 and 2.2 mg/L, respectively) can be divided into two segments in which first-order kinetics (with different reaction rate constants) prevails. When applying the proposed estimation-prediction algorithm at say $t = 48$ h, the period between $t = 15$ h and $t = 48$ h is used to derive the

regression line (Fig. 9) from which \hat{K} and $\hat{S}(0)$ are estimated [eq. (7)]. It is obvious that this straight line is only a first-order approximation, since it does not fit well the data between $t = 40$ h and $t = 120$ h. However, since the estimated parameters are used only to predict the residual DO consumption between $t = 48$ h and $t = 120$ h, the BOD_5 error is acceptable.

The summary of the test runs of this study (Table I) reveal that maximum BOD_5 prediction error at $t = 36$ h was 44% and at $t = 48$ h is 33%. The uncertainty is con-

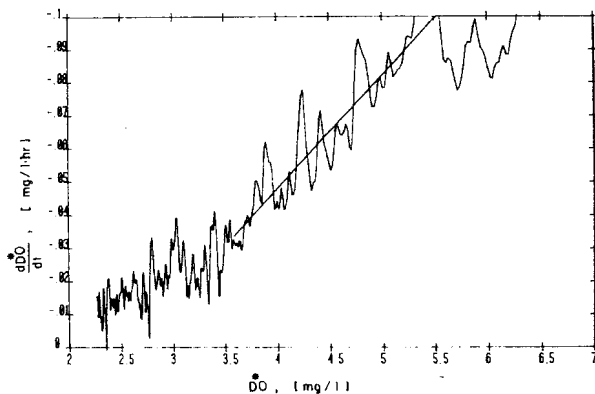


Figure 9. The relations between dDO/dt and DO in run Z104.

Table I. Measured and predicted BOD_5 of present study runs.

Sample No.	Date	Location	Sample rate (h^{-1})	Dilution	Temperature ($^{\circ}C$)	Measured amount (mg/L)	36-h prediction		48-h prediction	
							(mg/L)	Error (%)	(mg/L)	Error (%)
Z3	8/13/81	Hazerim	2	1:100	25 $^{\circ}C$	540	400	-25	450	-17
Z4	8/20/81	Hazerim	2	1:100	25 $^{\circ}C$	270	150	-44	180	-33
Z5	8/27/81	Hazerim	2	1:100	25 $^{\circ}C$	210	160	-24	170	-15
Z9	9/03/81	Beer-Sheva	2	1:1000	25 $^{\circ}C$	5000	6700	+34	5300	+6
Z103	5/06/82	Beer-Sheva	4	1:1000	30 $^{\circ}C$	7500	6600	-12	5900	-21
Z104	5/17/82	Beer-Sheva	4	1:1000	30 $^{\circ}C$	5200	4800	-8	4900	-6
Z105	5/27/82	Beer-Sheva	4	1:1000	30 $^{\circ}C$	7400	6800	-8	7000	-5
Z106	6/03/82	Beer-Sheva	4	1:1000	30 $^{\circ}C$	5570	4500	-19.2	4900	-12
Z107	6/10/82	Beer-Sheva	4	1:1000	30 $^{\circ}C$	5700	5100	-11	5300	-7

siderably smaller if data with sampling rate of 2 h^{-1} is excluded. It was found that the increased sampling rate improves the prediction markedly probably by providing more degrees of freedom to the least-squares fitting procedure. Reprocessing the higher sampling rate data by elimination of every other point and hence reducing the data point rate to two per hour increased the prediction error.

If the lower sampling rate data is excluded, the uncertainty range of the predicted BOD_5 is ca. 15% at $t = 36 \text{ h}$ and 10% at $t = 48 \text{ h}$.

CONCLUSIONS

The suggested estimation-prediction algorithm, which assumes a first-order kinetics implemented on an on-line computerized instrumentation system, could be used to accelerate BOD_5 analysis. This could increase the ability to control water treatment plants and in particular to alert the operators after only a short delay to possible malfunction of the plant and associated discharge networks. Such a system should include automatic sampling, discharging and aeration devices, as well as means for automatic dilution and reagent addition. All these devices could be under microcomputer control as only a small fraction of the realtime is needed for the estimation-prediction computation. A single microcomputer could in fact run a number of such experiments, some of which could be run over the full five-day period as on automatic validation of the estimation-prediction process. Since part of the irregularity in the observed DO consumption process could be due to sensor and instrumentation instability, a better understanding of the system-

atic and random errors of the analysis might enable one to improve the accuracy of BOD_5 prediction.

The financial support of the Ministry of Energy and Infrastructure of Israel is gratefully acknowledged.

References

1. T. D. Waite, "An evaluation of non specific measurements on organic pollutants in waters and waste waters," technical paper No. 34, Australian Water Resources Council, Sydney, Australia, 1978.
2. T. H. Tebbutt, *Principles of Water Quality Control* (Pergamon, Oxford, 1971).
3. *Standard methods for the examination of water and waste water*, American Public Health Association, New York, 1981.
4. N. K. Mullis and E. D. Schroder, *J. Water Pollut. Control Fed.*, **43**, 209 (1971).
5. R. M. Arthur, "BOD analyzer, general information," Arthur Bros. Co. Inc., Fond Du Lac, Wisconsin, 1971.
6. T. Viraraghavan, *J. Water Pollut. Control Fed.*, **48**, 2213 (1976).
7. S. Ben-Yaakov, R. Raviv, H. Guterman, A. Dayan, and B. Lazar, *Talanta*, **29**, 267 (1982).
8. T. H. Tebbutt and M. Berkum, *Water Res.*, **10**, 613 (1976).
9. L. G. Rich, *Environmental System Engineering* (McGraw-Hill, New York, 1973).
10. L. Van Davenee and W. W. Eckenfelder, *Water Res.*, **14**, 561 (1980).
11. D. Williamson, "Modelling and observation of microbial growth processes," Proceedings of the 5th IFAC, Paris, France, 1972.
12. S. M. Bozic, *Digital and Kalman Filtering* (Edward Arnold, London, 1979).
13. F. Schied, *Numerical Analysis* (McGraw-Hill, New York, 1968).
14. S. Ben-Yaakov, *Bamidge*, **31**, 69 (1979).
15. S. Ben-Yaakov, "Electrochemical instrumentation," in: *Marine Electrochemistry*, M. Whitefield and D. Jagner, Eds. (Wiley, New York, 1981), pp. 99-122.
16. P. R. Bevington, *Data Reduction and Error Analysis for Physical Science* (McGraw-Hill, New York, 1969).