Optimal design for dispersion experiment

ARTICLE *in* WATER RESEARCH · DECEMBER 2002

Impact Factor: 5.53 · DOI: 10.1016/S0043-1354(02)00170-7 · Source: PubMed

1 AUTHOR:

Jonah Agunwamba

University of Nigeria

59 PUBLICATIONS **166** CITATIONS

SEE PROFILE

PERGAMON

Water Research 36 (2002) 4570–4582

WATER RESEARCH

www.elsevier.com/locate/watres

Optimal design for dispersion experiment

J.C. Agunwamba

Department of Civil Engineering, University of Nigeria, Nsukka, Nigeria

Received 12 July 2001; received in revised form 29 January 2002; accepted 22 March 2002

Abstract

The dispersion coefficient (D) is a very important parameter for the management of water quality and pollution control. Conventional sampling methods are based on tracer studies. Sampling periods and intervals are often subjective and left to the choice of the experimenter. Sometimes several observations are made to determine D. Yet, making several observations does not necessarily ensure accurate estimation of the parameters. Sampling of tracer concentrations at poor times will result in inaccurate estimates of D and the flow velocity (u). Therefore, the main aim of the study in question was to investigate the optimal sampling times for conducting fixed position variable-time sampling for tracer concentration experiments. This objective was achieved by the minimization of the least square criterion and applying the method proposed by Box and Lucas. The relative efficiency of each experimental design is predicted before data collection and analysis and confidence regions plotted. Results, on optimal sampling times for experiments are presented and illustrated. The application of these findings will harmonize results and reduce cost and labour expended on dispersion experiments. © 2002 Published by Elsevier Science Ltd.

Keywords: Optimal design; Dispersion experiment; Least square criterion

1. Introduction

A dispersion coefficient (D) or its dimensionless form, dispersion number (d) is a very important parameter for the design of outfall structures. It is also useful for the determination of assimilatory and reaeration capacities of streams, and the general management of water quality and pollution control.

Several investigations have been done in modeling the dispersion of contaminants, prediction of the distribution of pollutant downstream from its point of discharge and evaluation of D from time-concentration curves [1–10].

In all the above investigations, tracer studies formed a very essential tool for model calibration and verification as well as characteristics studies. Although tracer studies have played a major role in the advancement of dispersion studies and pollution control generally, evaluation of experimental errors and improvement of sampling plans have not been studied much. A method of reducing the sampling period by sampling both at different times and positions along a channel has been proposed [1]. The present work is focused on the conventional fixed position sampling where only time is varied. Experiments are embarked upon without adequate design. This omission can result in excessive experimentation or inaccurate determination of D.

Reported variations [3] in D values are up to 100%. The assumption that sampling should be terminated 2.5Θ (where Θ is the detention time) after injection [7] is not always adhered to. Experiments have been terminated before 2Θ [11] while others have been terminated after $0.99, 2.66, 3.33$ and 4.44Θ [6]. Upto 101 data were used to compute a single value of d for Portimao primary facultative pond. The use of different sampling periods and intervals makes a realistic comparison of results from researchers and the interpretation of their results difficult. Besides, it is questionable whether one could have obtained a reasonably accurate value of D without resorting to using some excessive data. Of course reduction of the sample size will obviously result in reduction in cost and

E-mail address: epseelon@aol.com (J.C. Agunwamba).

0043-1354/02/$ - see front matter
PII: S0043-1354(02)00170-7

Nomenclature

A	channel or pond cross sectional area (m^2)
A_0	Hessian matrix, $H(\theta)$
d	dispersion number
D	dispersion coefficient (m^2/d)
E	largest difference between risks
f	dispersion equation evaluated at θ
F_x	upper (1-α) 100% point of the F-distribution
$H(\theta)$	Hessians matrix
l	dimension of ellipsoid
L	channel or pond length (m)
M	mass of tracer per channel volume
n	number of observations made
N	approximate Hessian
p	number of parameters
P	degree of freedom associated with S^2
s	sum of squares
S_c	critical level of sum of squares
S^*	minimum sum of squares
t	time
u	flow velocity
v	covariance matrix
x	derivative matrix
y_m	measured tracer concentration (mg/l)
y_p	predicted tracer concentration (mg/l)

Greek symbols

Δ	a measure of precision
Γ	gamma function
θ	vector representing u or D
Θ	detention time.

labour. In the light of the above problems, the objectives of the paper are:

(1) to investigate the adequacy of data currently collected during tracer studies and consequently the adequacy of the experimental strategies.
(2) to study carefully some typical examples of experiments conducted, identify poor designs and suggest ways in which greater efficiency can be obtained, and
(3) to suggest a sampling plan for which measurements will give the least error in the estimation of D and u.

2. Theoretical basis

2.1. Least square objective criterion

In this work, a one-dimensional dispersion equation is applied in the selection of the most efficient experimental design. The most efficient design is usually the one which gives the highest precision in parameter estimation. The precision of estimated parameters is often increased by using a larger number of observations. However, the quality of the estimated parameters depends also on the experimental design, that is, the actual choice of levels of the variables at which measurements are made.

The least square estimates of D and u are those values that minimize the sum of squares criterion

$$S = \sum_{i=1}^{n} [y_m - y_p]^2, \tag{1}$$

where

$$y_p = \frac{M}{A_0\sqrt{4\pi Dt}} \exp\ -\frac{(L-ut_i)^2}{4Dt_i} \tag{2}$$

y_m and y_p are measured and predicted tracer concentrations respectively, M is the mass of tracer per channel volume, A is the channel cross sectional area, L is the channel length and t is the time.

Unlike in linear equations where taking the derivatives of S partially with respect to D and u, equating each to zero and solving them simultaneously give optimum values of u and D, in the non-linear case, it is usually hard to solve explicitly. The location of the minimums often involves an iterative search of the parameter space using a digital computer. The initial estimates of u and D must be supplied. The different methods used in parameter estimation include linearization (Taylor series), steepest descent and Marquardt's Compromise [12].

Marquardt's compromise [13] represents a compromise between the linearization method and the steepest descent method and seems to work well in many circumstances and is thus a sensible practical choice. Because of the additional need to plot the contour, additional values of the sums of squares other than the minimum values are obtained.

2.2. Precision and reliability

It is not enough to obtain least square estimates of u and D, the reliability and precision of the estimates must be investigated. The precision of an estimate is represented by a confidence interval associated with an estimated parameter. A confidence region gives the joint precision of the parameters estimated simultaneously. A critical level of the sum of squares S_c may be calculated for delineating a confidence region:

$$S_c = S^* + S^2 P F_X(P, n-P) \tag{3}$$

in which S^* is the minimum sum of squares, P is the number of estimated parameters; $n-P$ is the degree of

freedom associated with the estimated variance S^2, and $F_X\ (p, n-P)$ is the upper $(1-\alpha)$ 100% point of the F-distribution with P and $n-P$ degrees of freedom [14,15]. For normally distributed experimental errors, an estimate of the experimental error variance, S^2, is expressed by

$$S^2 = \frac{S^*}{n-P}. \tag{4}$$

How much u and D can be varied from their optimal values were evaluated by the E-difference region defined by [12].

$$\theta = \theta \pm (2E/A), \tag{5}$$

where θ is a vector, in this case, representing either u or D which is the equation of an one-dimensional ellipsoid whose volume is $(2E\pi)^{1/2} \det^{-1/2} A/\Gamma(L/_2 + 1)$; A is the Hessian matrix $H(\theta)$ of the function of the second partial derivatives; that is

$$H(\theta) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial u^2} & \dfrac{\partial^2 f}{\partial u \partial D} \\ \dfrac{\partial^2 f}{\partial u \partial D} & \dfrac{\partial^2 f}{\partial D^2} \end{pmatrix} \tag{6}$$

f represents the dispersion equation evaluated at optimal values of $\theta(u, D)$, and E is the largest difference between risks we are willing to consider insignificant.In order to estimate the covariance matrix ($\boldsymbol{V}$) in the case of single least squares, the equation

$$\boldsymbol{V} = S^2 \left[\sum_{i=1}^{n} (\partial f_i/\partial\theta)(\partial f_i/\partial\theta)^{\mathrm{T}} \right]^{-1} \tag{7}$$

is used where S^2 is the variance determined from Eq. (4) and $(\partial f_i/\partial\theta)^{\mathrm{T}}$ is the transpose matrix of $\partial f_i/\partial\theta$.

Eq. (7) may be reduced to

$$\boldsymbol{V} = 2S^2 \boldsymbol{N}^{-1} \tag{8}$$

in which $\boldsymbol{N}$ is the approximate Hessian

$$\boldsymbol{N} = 2 \sum_{i=1}^{n} (\partial f_i/\partial\alpha)(\partial f_i/\partial B) \quad (\alpha, \beta = 1, 2) \tag{9}$$

in vectorial form evaluated at the optimum values of u and D.

3. Experimental design

The design is aimed at determining the sampling times (for instance, t_1, t_2, t_3, and so on) for the tracer, which will yield the best estimates for u and D. This was done by maximising the determinant of the derivative matrix, X which Berthouex and Hunter [14] defined as

$$X = (X_{ij}) = \begin{pmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ & & \vdots & \\ X_{1n} & X_{2n} & \dots & X_{pn} \end{pmatrix} \tag{10}$$

in which p is the number of parameters, n is the number of observations made and

$$X_{ij} = \partial f(\theta, t)/\partial\theta_i, \quad i = 1, 2 \ldots, p,\ j = 1, 2, \ldots, n. \tag{11}$$

For $p = 2$ and $n = 2$

$$\boldsymbol{X} = \begin{vmatrix} X_{11} & X_{12} \\ X_{12} & X_{22} \end{vmatrix}, \tag{12}$$

$$\begin{pmatrix} \dfrac{a_1}{2D}(L - ut_1)\exp(-b_1) & \dfrac{-a_1}{2D} + \dfrac{a_1(L-ut_1)^2}{4D^2 t_1}\exp(-b_1) \\ \dfrac{a_2}{2D}(L - ut_2)\exp(-b_2) & \dfrac{-a_2}{2D} + \dfrac{a_2(L-ut_2)^2}{4D^2 t_2}\exp(-b_2) \end{pmatrix}, \tag{13}$$

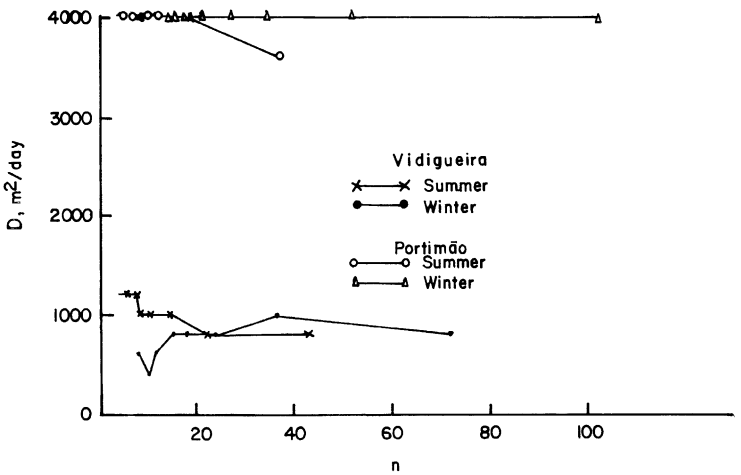

Fig. 1. Variation of coefficient of dispersion D, with the number of experimental points, n for Vidigueira and Portimão with the peak concentrations included.

where

$$a_1 = \frac{1}{\sqrt{4\pi D t_1}}, \quad a_2 = \frac{1}{\sqrt{4\pi D t_2}}, \quad b_1 = \frac{(L-ut_1)^2}{4Dt_1} \tag{14}$$

and

$$b_2 = \frac{(L-ut_2)^2}{4Dt_2}.$$

The determinant of the matrix in Eq. (13), Δ is obtained as

$$\Delta = \frac{a_1(L-ut_1)}{2D}\exp(-b_1+b_2)\left(\frac{-a_2}{2D}+\frac{a_2(L-ut_2)^2}{4D^2t_2}\right) - \frac{a_2}{2D}(L-ut_2)\left(\frac{-a_1}{2D}+\frac{a_1(L-ut_1)^2}{4D^2t_1}\right)\exp(-b_1+b_2). \tag{15}$$

The analytical solution is found by setting the partial of the derivatives of Δ with respect to t_1 and t_2 to zero and finding the simultaneous solution of the two resulting equations. For a two-parameter model with n observations the value of the determinant is

$$\Delta = \sum_{i=1}^{n}(x_{ij})^2 \sum_{j=1}^{n}(x_{2j})^2 - \left(\sum_{i=1}^{n} x_{1j}\, x_{2j}\right)^2. \tag{16}$$

Table 1
Variations of flow velocity (u), dispersion coefficient (D) minimum least squares sum (S) and critical least squares sum (S_c) with the number of data points for the four experiments (with peak)

Experiment	Estimated parameters			Sum of squares		Number of data points
	Skip	u (m/d)	D (m^2/d)	S	S_c	n
	0	10	800	384	445	43
	1	11	800	227	306	22
	2	11	1000	162	257	15
1	3	12	1000	141	275	11
	4	12	1000	144	339	9
	5	15	1200	116	315	8
	6	16	1200	111	368	7
	7	19	1200	111	496	6
	0	12	800	506	551	72
	1	12	1000	285	340	36
	2	11	800	192	252	24
2	3	11	800	175	254	18
	4	12	800	168	267	15
	5	11	600	149	271	12
	6	12	400	139	271	11
	7	10	600	131	308	9
	0	23	3600	92	109	36
	1	25	4000	48	70	18
	2	24	4000	38	68	12
3	3	24	4000	31	73	9
	4	23	4000	29	79	8
	5	30	4000	31	136	6
	6	25	4000	34	151	6
	7	29	4000	29	210	5
	0	30	4000	1550	1647	101
	1	30	4000	768	868	51
	2	30	4000	585	706	34
4	3	30	4000	567	728	26
	4	29	4000	370	507	21
	5	30	4000	339	505	17
	6	30	4000	321	509	15
	7	29	4000	237	409	13

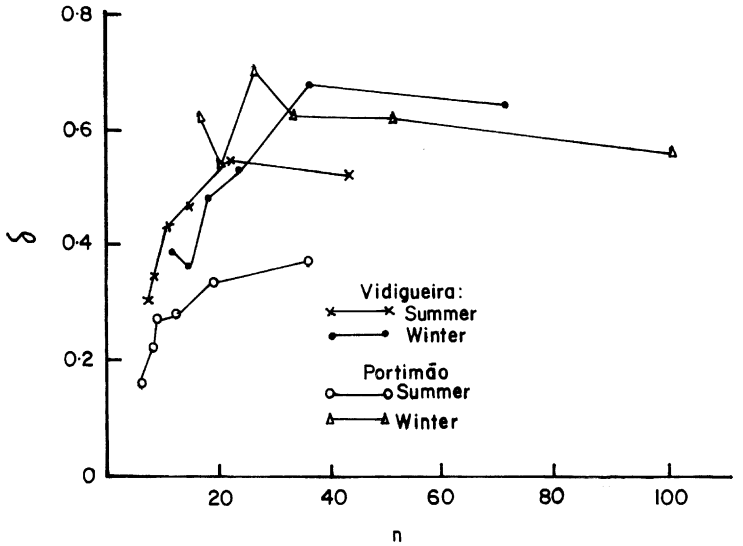

Fig. 2. Variation of dispersion number, δ with the number of data points, n for Vidigueira and Portimão ponds with the peak concentrations included.

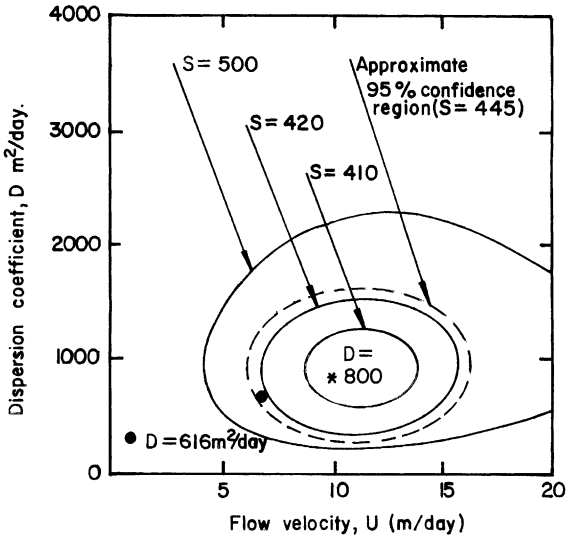

Fig. 3. Least square contours for experiment runs 1 with high U and D ranges.

Δ is a measure of the precision of design. $\Delta^{1/2}$ is inversely proportional to the size of the confidence region. The larger $\Delta^{1/2}$ is the greater the precision.

3.1. Data source

Two sets of data were used for the illustration of an optimal experimental design: laboratory and full scale waste stabilization pond data. The laboratory data were based on tracer experiments conducted by the author in a 14 m channel in the Department of Civil Engineering Hydraulic laboratory already published in a different paper [1]. The full-scale pond data were from tracer experiments conducted in Portugal [6]. These experimental results will be examined with respect to the adequacy of their design, precision and reliability. Alternative designs involving a small number of observations shall be investigated.

4. Results and discussion

4.1. Variation of d with number of data points (n)

Fig. 1 and Table 1 show the variation of D with the number of experiments when the peak concentrations are included in the determination. The data were obtained from the literature [6]. In all cases the same optimal or close-to-optimal values can be obtained by using a smaller number of experimental points. For instance, in experiment run 2, the optimal pair of values are $u = 12$ m/d and $D = 800\,m^2/d$ for $n = 72$. This same pair of values was obtained when the number of observations is 15. These results emphasize the possibility of obtaining the optimal values with smaller experimental data (Table 1). However, the corresponding precision of the estimates is also very important.

Fig. 2 above shows the same variation as in Fig. 1 but with D replaced with the dispersion number, $d(= D/ul)$.

The dispersion number appears to be more sensitive to u than D is to n. This is attributable to the effect of u in the formula for d computation.

4.2. Contours

The contours of the sum of least squares for the estimation of u and D are shown in Fig. 3 for experiment run 1. The confidence region is small. Within 95% confidence the true values of u and D are included within this region. As expected the contours are not ellipsoidal but irregular and 'banana-shaped'. The optimum estimates are $u = 10\,\mathrm{m/d}$ and $D = 800\,\mathrm{m^2/d}$. The contours are approximately well-rounded ellipses. Hence, the surface is well conditioned. The axes of the ellipses are parallel to the u and D axes, hence the value which makes S a minimum has no dependence on D. That is, specific information about D which fixed its value would not alter the least squares estimate u. Since the longer axis is parallel to D axis, D is ill-determined whereas u is well determined. The smaller scale used on the u-axis seems to give the impression in Fig. 3 that the reverse is the case. That u is well conditioned unlike D which is made clearer in Fig. 4. For all the ranges of u and D (0–80 and 0–4000$\mathrm{m^2/d}$, respectively) considered, there were two global minimums. At higher values of u and D the minimum occurred at $(u, D) = (10, 800)$ (Fig. 3), at low values at $(u, D) = (0.5, 1.49)$ (Fig. 4). The critical least square criterion (S_c) and the least square criterion (S) vary significantly with n. The values of n where $S_c > S$ will obviously yield a low precision of the estimates. $S_c < S$ does not necessarily imply that the values of n and selection of those corresponding sampling times will give a high precision since the criterion is only approximate for non-linear functions. However, the results also confirm that an estimation of D and u with reduced data requirement is possible.

Table 2 shows the computation of S and S_c without the peak values. The absence of the peak values does not have much effect on the optimal values. A small sampling interval, small differences between the peak and neighbouring concentrations as well as multiple peaks were responsible. In fact, because of the extent of short-circuiting observed, the removal of the peak reduced the variance as Tables 3 and 4 show. Even though u and D can be estimated with smaller data points, it is necessary to examine the precision of the estimates. The precision and reliability are investigated by computing the E-difference using relationships already expressed in Eqs. (3)–(9).

Consider the optimal data for experiment run 1 for $n = 43$ (Table 1).

$$\sigma^2 = 384/(43 - 2) = 9.3659$$

therefore

$$\sigma = 3.0604$$

$$V = 17,097.714\begin{pmatrix} 8.605 \times 10^{-5} & 7.5109 \times 10^{-3} \\ 7.510875 \times 10^{-3} & 7.02141 \end{pmatrix}.$$

The standard deviations of u and D are $\sigma_u = (8.6051 \times 10^{-5} \times 17,097.714)^{1/2} = 1.21296$ and $\sigma_D = (7.02141 \times 17,097.714)^{1/2} = 346.4824$. The correlation between the estimates is

$$\frac{7.510875x10^{-3} \times 17,097.714}{1.21296 \times 346.4824} = 0.3056.$$

Since the variance of D is high it is ill-determined [12].

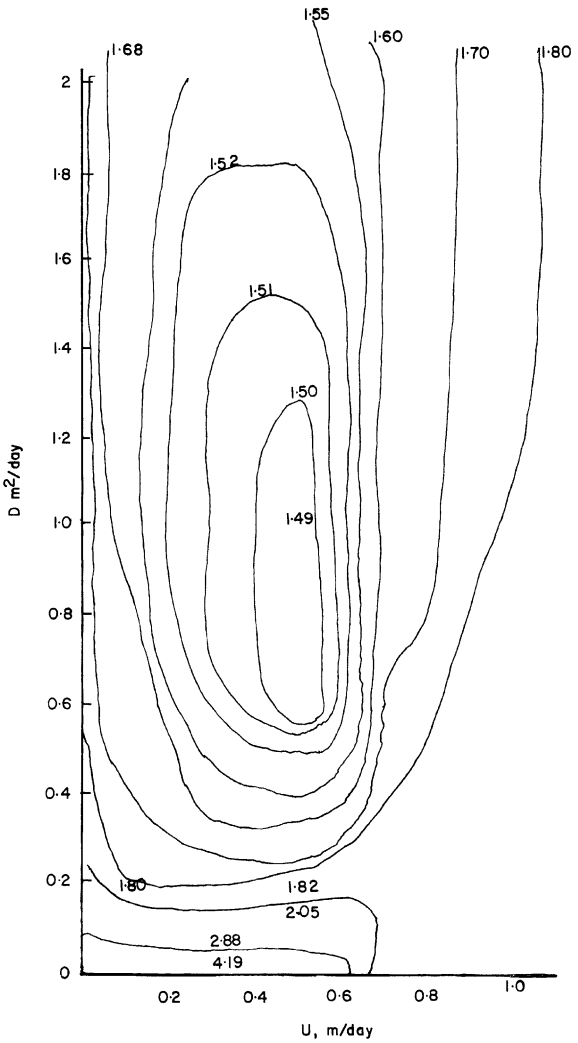

Fig. 4. Least square contours for experiment runs 1 with low U and D ranges.

Table 2
Variations of flow velocity (u), dispersion coefficient (D) minimum least squares sum (S) and critical least squares sum (S_c) with the number of data points for the four experiments (without peak)

Experiment	Estimated parameters			Sum of squares		Number of data points
	Skip	u	D	S	S_c	n
	0	10	800	294	342	42
	1	10	800	137	188	21
	2	9	800	67	111	14
1	3	9	600	42	89	10
	4	9	400	32	87	8
	5	8	600	21	69	7
	6	8	400	17	62	6
	7	8	400	17	124	5
	0	12	1000	438	478	71
	1	13	1000	193	232	35
	2	12	800	85	113	23
2	3	13	1000	63	94	17
	4	13	800	55	90	14
	5	13	800	33	64	11
	6	13	600	20	42	10
	7	11	600	11	29	8
	0	23	3800	77	93	35
	1	24	3600	40	60	17
	2	24	3400	24	47	11
3	3	21	3400	16	45	8
	4	18	3400	15	49	7
	5	25	3400	18	131	5
	6	19	3400	18	136	5
	7	22	3400	16	318	4
	0	30	4000	1440	1541	100
	1	30	4000	635	720	50
	2	30	4000	473	574	33
4	3	30	4000	186	241	25
	4	30	4000	357	498	20
	5	30	4000	224	344	16
	6	30	4000	187	308	14
	7	30	4000	87	158	12

Using the E-difference formula (Eq. (5)), $(2E/A) = 1.067$ and 304.9 for u and D respectively. Hence, the ranges of u and D are $8.933 \leqslant u \leqslant 11.067$ and $495.1 \leqslant D \leqslant 1104.9$, respectively for $E = 4$.

The design locus of dispersion for the four experiments is shown in Fig. 5. The experimental observations to be made must be chosen from the locus of points on a curved line in the (x_1, x_2) space where $x_1 = \partial f/\partial u$ and $x_2 = \partial f/\partial D$. The function x_2 has been plotted against x_1 for the four experiments at the corresponding optimal values of u and D. As t goes from 0 to ∞ these pairs trace out the curve shown in Fig. 5, called the *design locus* which give the maximum area at the optimum sampling times. The dotted lines in experiment run 3 shows the points as $t_1 = 3.521$ days and $t_2 = 14.927$ days. Additional points are at $t = 0$ and ∞. These are also the values found by straightforward maximization of the design criteria except the point $t = 64.99$ which is as good as $t = \infty$ considering the fact that most of the tracer would have almost completely decayed at that point.

4.3. Comparison between least square and levenspiel D and u

D and u are indicated in Fig. 6. Estimation of D and u has always been a problem. None of the methods proposed so far yields best values of D and u because they are influenced to a greater or less degree by deviations of the observed data from the model [10]. Methods of estimating D from $t-c$ curves are unreliable at best and highly inaccurate at worst. The principle of

Table 3
Ranges of estimated parameters and their corresponding correlation coefficients and standard deviations for $E = 4$ with peak concentrations included

Experiment	Ranges of estimated parameters		Correlation
	u	D	
1	10 ± 1.067 (1.2130)[a]	800 ± 304.9 (346.5)	0.3056
2	12 ± 0.373 (0.515)	1000 ± 178.0 (245.7)	0.2264
3	23 ± 1.377 (1.216)	3600 ± 988.6 (873.2)	0.3646
4	30 ± 0.381 (0.569)	400 ± 1092.7 (2162.3)	−0.0203

[a] Figures in parenthesis are the standard deviations.

Table 4
Ranges of estimated parameters and their corresponding correlation coefficients and standard deviations for $E = 4$ without peak concentrations included

Experiment	Ranges of estimated parameters		Correlation
	u	D	
1	10 ± 0.782 (1.069)[a]	800 ± 91.6 (125.2)	0.1275
2	12 ± 0.394 (0.510)	1000 ± 178.0 (230.8)	0.2370
3	23 ± 1.428 (1.167)	3600 ± 992.6 (810.9)	0.3550
4	30 ± 0.387 (0.741)	4000 ± 1092.7 (2094.8)	−0.0106

[a] Figures in parentheses are the standard deviations.

least square estimation of non-linear parameters must be used to provide acceptable accuracy. The optimum D and u vary remarkably from values computed by the Levenspiel and Smith method as shown in Fig. 6.

4.4. Determination of the optimum sampling times

The two equations for t_1 and t_2 are similar and solving them will give the same values. Solving the equations $d\Delta/dt = 0$, we obtain

$$T = L/u,\ t = 0,\ t = \infty \text{ and } L_4 - 2L^3ut + 2Lu^3t^3 - u^4t^4 + 2D^2t^2 - 7DtL^2 + 5u^2t^2D + 2LuDt^2 = 0. \quad (17)$$

Dividing by L^4 right through we obtain

$$1 - \frac{2ut}{L} + \frac{2u^3t^3}{L^3} - \frac{u^4t^4}{L^4} + \frac{2D^2t^2}{L^4} - \frac{7Dt}{L^2} + \frac{5u^2t^3D}{L^4} + \frac{2uDt^2}{L^3} = 0. \quad (18)$$

The equations can be split into two in order to examine the roots.

That is

$$y_1 = -1 + P_1t^4 - P_2t^2 \quad (19)$$

and

$$y_2 = M_1t^3 - M_2t, \quad (20)$$

where

$$P_1 = u^4/L^4, \quad P_2 = 2D^2/L^4 + 2uD/L^3, \quad M_1 = 2u^3/L^3 + 5u^2D/L^4, \quad M_2 = 2u/L + 7D/L^2. \quad (21)$$

The intersection of the two curves will give the roots of the original equation. The number of real roots will depend on the coefficients of the terms. When the curves are sketched and all possible intersections identified, two cases emerge. There are 3 real roots if $y_2\sqrt{M_2/3M_1} < 1$ and only one real root if $y_2\sqrt{M_2/3M_1} > 1$. The real roots occur after $t > [-P_2 + (P_2^2 + 4P_1)^{1/2}]^{1/2}/(-2P_1)^{1/2}$, between $t = (P_2/2P_1)^{1/2}$ and $[M_2/M_1]^{1/2}$, and between $t = 0$ and $\sqrt{M_2/3M_1}$).

Depending on the values of M_1 and M_2, $y_2\sqrt{M_2/3M_1}$ is greater or less than 1. For instance, for $D = 3600\,m^2/d$, $u = 21\,m/d$ and $L = 350.2\,m$, $M_1 = 1.1997 \times 10^{-3}$, $M_2 = 0.3368$. Hence, $y_2\sqrt{M_2/3M_1} < 1$. Hence, there will be 3 real roots (3.373, 14.139, 64.990). However, for very low values of d ($D = 0.015\,m^2/s$, $u = 0.35\,m/s$ and $L = 14\,m$), $y_2\sqrt{M_2/3M_1} > 1$ and there will be only one root (21.2846 s). It is necessary then to ensure that samples are collected at least once between each of the above three sets of ranges in order to maximize the precision. Additional values may be collected at $t = L/\theta$ and for as long a time as possible after the injection of tracer ($t \to \infty$).

The roots of this equation will give the required values of optimal times to conduct the experiment.

Again, experiment run 3 is used for an illustration. With the optimum values of u and $D = 23$ and $3600\,m^2/d$, respectively and $L = 350.2\,m$, the Newton–Raphson's method gives –17.9332, 3.273, 14.1387 and 64.990 days within an error of $\pm 10^{-3}$. Other values are $L/u, t = 0$ and ∞.

4.5. Experimental designs

Two cases are considered and they are based on data collected from a laboratory channel and a field pond respectively. By choosing optimum sampling times the data requirement can be reduced. The laboratory case is given in Table 5 and Fig. 7 involving 6 different

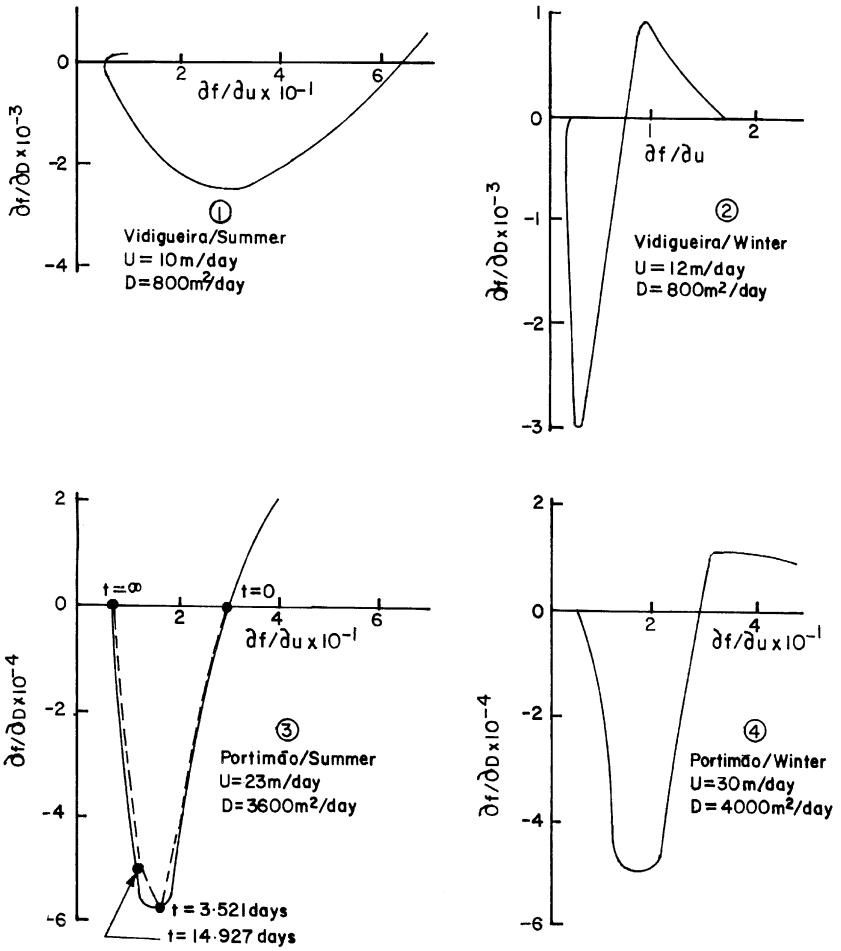

Fig. 5. Design locus of dispersion for Vidigueira and Portimão waste stabilization ponds for experiments 1–4.

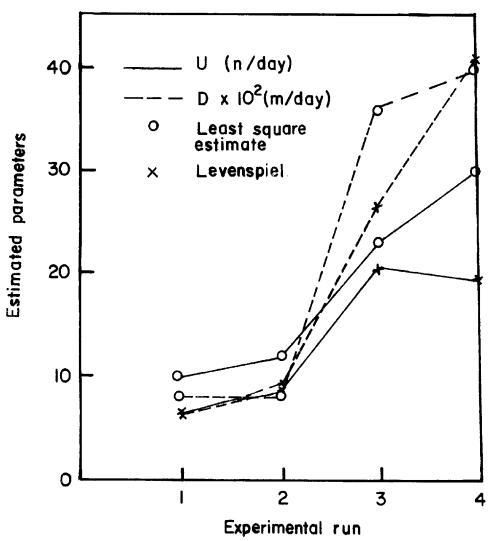

Fig. 6. Comparison of least square and Levenspiel estimates of U and D for the four experiments runs.

Table 5
Evaluation of design determinant Δ for six typical dispersion coefficient and flow velocity estimation

Typical design (Fig. 6)	Number of points (n)	$D = 0.015\,m^2/s$, $u = 0.35\,m/s$, $L = 14\,m$	
		Δ	Remark
a	12	210,515.1	Excluding some tail points
b	12	285,220.4	Excluding some initial points
c	12	177,244.1	Without points around L/u
d	10	351,167.8	Cluster around roots
e	10	2,794,284.1	Replicate
f	15	359,303.7	All 15 observations

experimental designs. For this experiment with $u = 0.35\,m/s$, $D = 0.015\,m^2/s$ and $L = 14\,m$, the roots of the equation are $t = 0, L/u$ (40 s) and ∞. The precision of estimation of u and D using each of the designs depends on the values of Δ.

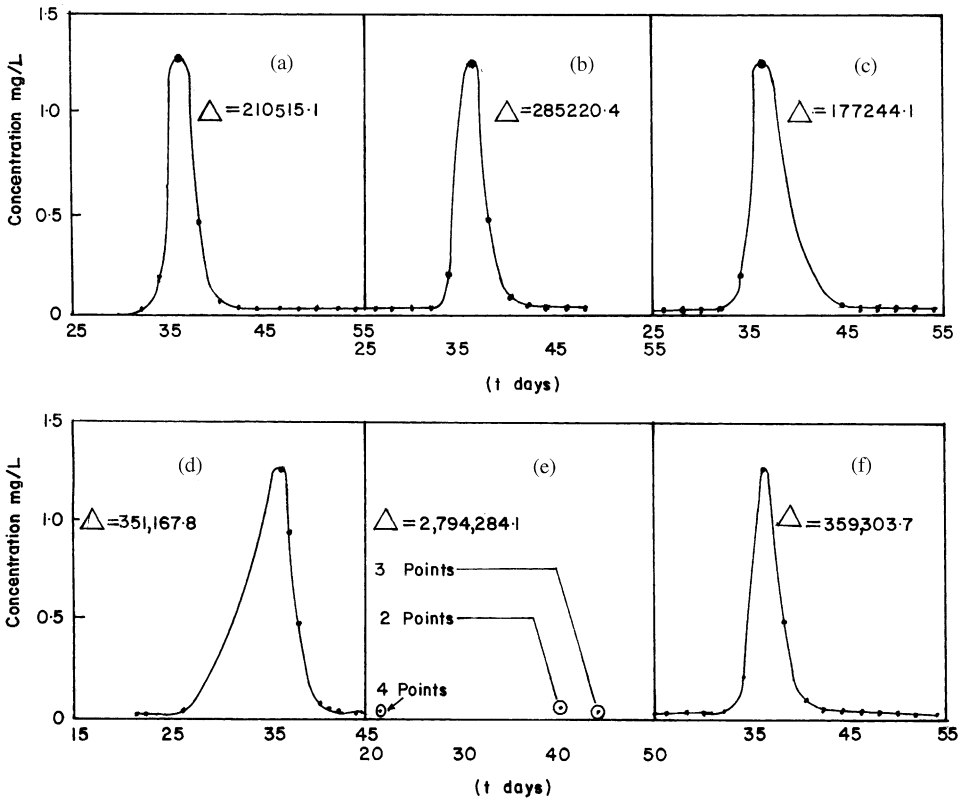

Fig. 7. Typical dispersion experimental designs with values of Δ for $D = 0.015\,m^2/s$, $U = 0.35\,m/s$ and $L = 14\,m$ (laboratory channel).

Table 6
Evaluation of design determinant Δ for six typical dispersion coefficients and flow velocity estimation

Typical design (Figs. 7 and 8)	Number of points (n)	$D = 3600\,m^2/day$, $u = 23\,m/day$, $L = 350.2\,m$	
		Δ ($\times 10^{-20}$)	Remark
a	36	4.4634	All observations
b	31	3.2542	Excluding 5 tail points
c	31	0.71425	Excluding 5 initial points
d	31	0.0088751	Excluding sampling at $t = 10$–15 days
e	18	1.9710	Starting from first non-zero point skip 1 point
f	18	7.5221	Clustering around roots—8 experimental points around 3.221; 6 around 15.221 (or 14.139) and 4 from $t = 31$–37 s

Generally, clustering all observations before the peak alone or after the peak alone is not a good design. In each case the value of $(\Sigma x_{ij}x_{2j})^2$ is maximized, thereby reducing the overall efficiency. Designs *a* and *b* show that removal of tail points (a) has less significant effect on Δ than when earlier readings were removed (b). This result is very crucial especially since conventionally much emphasis is placed on the long tail than at the beginning. Experiments should therefore be conducted to collect points before the peak. This calls for carefulness and readiness to start sampling before the peak is quickly reached. However, cutting off all the tail points entirely will mean that the path of the curve at the lower end will be uncertain. Such points are necessary to reduce the uncertainty in the path of the curve.

In design *c*, removal of values at $t = 38$, 40 and 42 s produces a very significant reduction in Δ by 182,060. The inclusion of points around L/u (40 s) is crucial for maximization of Δ. Taking readings at $t = L/\theta$ helps to maximize the value of Σx_{ij}^2 while minimizing $\Sigma(x_{ij}x_{2j})^2$.

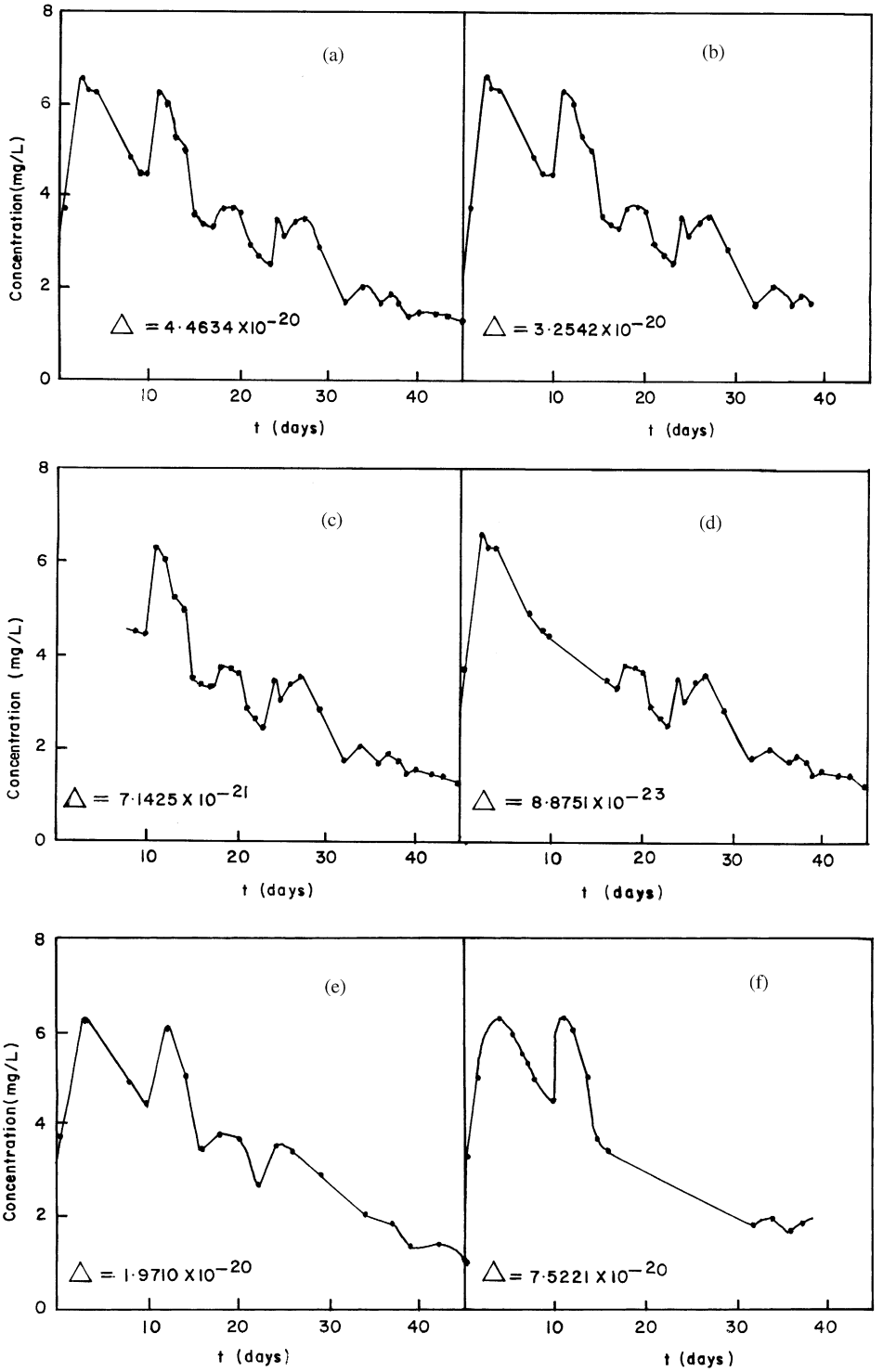

Fig. 8. Typical dispersion experimental designs with values of Δ for $D = 3600\,m^2/d$, $U = 23\,m/d$ and $L = 350.2\,m$ (Portimão WSP).

Well-positioned 10 points clustered around $t = 21.2846, L/\theta$, 36, 37, 40 and 42 with other filling points at $t = 26$, 41 and 44 gave a very high Δ. The 10 points yielded almost the same value with the 15 points. Design (d) illustrates that provided the experimental design is efficient a precise estimation of parameters can be

obtained without a large number of observations. Design (f) is the whole set of 15 points obtained experimentally [1] while design (e) is replicated. A number of 4, 2, 3 and 1 readings were taken at $t =$ 21.2846, 40, 44 and 50 s, respectively. The values of $\Delta = 2,794,284$. This implies that $\Delta^{1/2}(= 1671.61)$ is three times higher than using 15 unreplicated samplings ($\Delta^{1/2} = 599.42$). A replicated design will produce more precise estimates with fewer observations.

Similar results were obtained in experiment run 3 in a full-scale waste stabilization pond (Table 6 and Figs. 8a and b). Designs b, c and d have 31 points each. The removal of five points at the tail end resulted in a lower reduction of Δ than removal of the initial points. Again, this emphasizes the importance of sampling early within the diffused period to establish the trend of the rising curve. Sampling without including $t = 10$ to 15 days gave a value of $\Delta = 8.8751 \times 10^{-17}$ (design d). It is very essential to sample around L/θ ($= 15.226$ days).

Designs (e) and (f) each has 18 points. In design (e), retaining the first non-zero point, alternative points are omitted. The value of $\Delta = 1.97104 \times 10^{-20}$. By clustering 9 readings from 0.1 to 8.0, 5 readings around 15.226 (i.e. L/θ) and taking the next 4 from $t = 31$ to 37 days), $\Delta = 7.5221 \times 10^{-20}$. While the first 9 readings including the optimal $t = 3.273$ days fix the rising portion of the curve the last points are meant to fix the lower portion of the curve.

Well-selected 18 points of design e yielded a better precision than making 36 observations (design a) or skipping some observations alternately within the period (design e). Even though one of the roots is 64.990 days, stopping the sampling about 2θ (30.4522 days) as was done here will give a high precision. However, the actual period to stop sampling shall be treated in the next paper.

5. Conclusion

The paper has investigated an experimental design for the estimation of dispersion coefficient and flow velocity. The Box and Lucas [15] method was used to examine design of experiments to achieve high precision. Sampling can be predicted before the experiment and this greatly facilitates the steps in obtaining good estimates of u and D. Making a great number of observations does not necessarily produce corresponding high precision estimates that are commensurate with work expended. The following conclusions can be drawn from the work, both from the field pond and laboratory data analysed.

1. As expected the least square contours are not ellipsoidal but tend to be irregular and often "banana-shaped," with two global minimums, one at high u and D values and the other at low values of d.
2. With respect to the two parameters u and D, u appears more well conditioned than D at low velocity. The reverse is the case at a high flow velocity.
3. The number of observations currently made in some experiments can be whittled down while still obtaining good estimates at high precision by clustering the sampling around L/θ and other optimal times given by the roots of $d\Delta/dt = 0$. There could be three or one real roots depending on the values of D, u and L. Eighteen (18) properly located points were found to be essentially as good as 36 points. Obviously, making a small number of observations has some savings in terms of cost and energy.
4. Several observations should be made before the peak to fix the trend of the rising curve accurately. Fixing the rising curve is more critical for high precision than fixing the tail end.

However, detailed analysis should be conducted with more data to investigate the applicability of the above method in planning tracer studies for natural water systems.

References

[1] Agunwamba JC. Reduction of sampling time in tracer studies. Water Environ Res 1997;63:343–9.
[2] Elder JW. The dispersion of marked fluid in turbulent shear flow. J Fluid Mech 1959;5(4):544–60.
[3] Fischer HB. The mechanics of dispersion in streams. J Hydrol Div ASCE 1967;93(6):187–216.
[4] Harris EK. A new statistical approach to the one-dimensional diffusion model. Int J Air Water Pollut 1963;7:799.
[5] Levenspiel O, Smith WK. Notes on the diffusion type model for the longitudinal mixing of fluids in flow. Chem Eng Sci 1957;6:227–33.
[6] Marecos do Monte MHF, Mara DD. The hydraulic performance of waste stabilization ponds in Portugal. Water Sci Tech 1987;19(12):219–27.
[7] Ojiako GU. An integrated study of tracer and solids particle dispersion. IE(I) J EN 1988;6a:25–31.
[8] Parker FL. Eddy diffusion in reservoirs and pipelines. J Hydrol Div ASCE 1961;87(HY3):151–71.
[9] Taylor GI. The dispersion of matter in turbulent flow through flow through a pipe. Proc Roy Soc London A 1954;223:446–68.
[10] Thackston EL, Hays JR, Krenkel PA. Least squares estimation of mixing coefficients. J Sannit Eng Div Am Soc Civ Engrs 1967;93:47–58.
[11] Olanrewaju MO, Ogunrombi JA. Improving facultative ponds hydraulics with baffles. Nigerian Eng 1992;27(2):9–20.

[12] Bard Y. Non linear parameter estimation. New York: Academic Press, 1974. pp. 170–217.
[13] Marquardt DL. An algorithm for least squares estimation of non-linear parameters. J Soc Ind Appl Math 1963;2:431–41.
[14] Berthouex PM, Hunter WG. Problems associated with planning BOD experiments. J Sannit Eng Div ASCE 1971;97(SA3):333–44.
[15] Box GEP, Lucas HL. Design of experiments in non-linear situations. Biometrika 1959;46:77–90.