- 22. Watts, J. D., Rittby, M. and Bartlett, R. J., J. Am. Chem. Soc., 1989, 111, 4155.
- 23. Haque, M. and Kaldor, U., Chem. Phys. Lett., 1985, 117, 347;
- 24. Haque, M. and Kaldor, U., Chem. Phys. Lett., 1985, 120, 261,
- 25. Kaldor, U., Int. J. Quantum Chemistry Symp., 1986, 20, 445.
- 26. Kaldor, U., J. Comput. Chem., 1987, 8, 448.
- 27. Kroto, H. W., Matti, G. Y., Suffolk, R. J., Watts, J. D., Rittby, M. and Bartlett, R.J., J. Am. Chem. Soc., 1990, 112, 3779.
- Sinha, D., Mukhopadhyay, S. K., Prasad, M. D. and Mukherjee,
 D., Chem. Phys. Lett., 1986, 125, 213; Sinha, D., Mukhopadhyay,
 S. K., Prasad, M. D. and Mukherjee, D., Chem. Phys. Lett., 1986, 129, 369.
- 29. Koch, S. and Mukherjee, D., Chem. Phys. Lett., 1988, 145, 321;
- 30. Bartlett, R. J., J. Phys. Chem., 1989, 93, 1697; Kvasnicka, V., Laurine, V. and Biskupic, S., Phys. Rep., 1982, 90, 159.
- 31. Mukherjee, D. and Pal, S., Adv. Quant. Chem., 1989, 90, 291.
- 32. Martensson-Pendrill, A. and Ynnerman, Y., Physica Scr., 1990, 41, 329.
- 33. Noga, J. and Urban, M., Theor. Chim. Acta, 1988, 73, 291.
- 34. Diercksen, G. H. F., Roos, B. E. and Sadlej, A. J., Chem. Phys., 1981, 59, 29.
- 35. Koch, H. and Jørgensen, P., J. Chem. Phys., 1990, 93, 3333.
- 36. Salter, E. A., Trucks, G. W. and Bartlett, R. J., J. Chem. Phys., 1989, 90, 1752.
- 37. Monkhorst, H. J., Int. J. Quantum Chem., 1977, S11, 421.
- 38. Sekino, H. and Bartlett, R. J., Int. J. Quantum Chem., 1984, S18, 255; Sekino, H. and Bartlett, R. J., J. Chem. Phys., 1986, 84, 2726.
- 39a. Pal, S., Theor. Chim. Acta, 1985, 68, 379; Pal, S., Phys. Rev., 1986, A33, 2240.
- 39b Pal, S. and Ghose, K. B., Phys. Rev., 1992, A45, 1518.
- 40. Ghose, K. B., Nair, P. G. and Pal, S. (Submitted to Chem. Phys. Lett.)
- 41. Pal, S., Theor. Chim. Acta, 1984, 66, 151.

- 42. Pal, S., Phys. Rev., 1990, A42, 4385.
- 43. Pal, S., Int. J. Quantum Chem., 1992, 41, 443.
- 44. Pal, S., Phys. Rev., 1989, A39, 39.
- 45. Pal, S., Prasad, M. D. and Mukherjee, D., Theor. Chim. Acta., 1984, 66, 311.
- 46. Pal, S., Prasad, M. D. and Mukherjee, D., Theor. Chim. Acta., 1983, 62, 523.
- 47. Ghose, K. B. and Pal, S., Chem. Phys. Lett., 1991, 187, 637.
- 48. Arponen, J., Ann. Phys., 1983, 151, 311.
- 49. Arponen, J., Bishop, R. F. and Pajanne, E., Phys. Rev., 1987, A36, 2519; Phys. Rev., 1987, A36, 2539;
- 50. Arponen, J., Bishop, R. F. and Pajanne, E., Lecture Notes in Chemistry (ed. Mukherjee, D.), Springer Verlag, 1989, vol. 50, p. 79.
- 51. Pal, S., Theor. Chim. Acta, 1984, 66, 207.
- 52. Kutzelnig, W., Mod. Theor. Chem., (ed. Schaeser, H. F. III), Plenum, N. Y., 1977, vol. 4, p. 129.
- 53. Bartlett, R. J. and Noga, J., Chem. Phys. Lett., 1989, 150, 29.
- 54. Bartlett, R. J., Kucharski, S. A. and Noga, J., Chem. Phys. Lett., 1989, 155, 133.
- 55. Pal, S., Phys. Rev., 1986, A34, 2682.
- 56. Ghose, K. B. and Pal, S., Phys. Rev., 1987, A36, 1539.
- 57. Pal, S. and Bartlett, R. J., Theor. Chim. Acta (To be published)
- 58. Salter, E. A., Sekino, H. and Bartlett, R. J., J. Chem. Phys., 1987, 87, 502.
- Meissner, L. and Bartlett, R. J., J. Chem. Phys., 1990, 92, 561;
 Meissner, L., Kucharski, S. A. and Bartlett, R. J., J. Chem. Phys., 1989, 91, 6187;
 Mukhopadhyay, D. and Mukherjee, D., Chem. Phys. Lett., 1989, 163, 171.
- 60. Dunning, T. H., J. Chem. Phys., 1970, 53, 2823.
- 61. Mounter, J. S. and Klemperer, W., J. Chem. Phys., 1970, 52, 6033.
- 62. Werner, H. J. and Meyer, W., Mol. Phys., 1976, 31, 855.
- 63. Mounter, J. S., J. Chem. Phys., 1972, 56, 5409.

RESEARCH COMMUNICATIONS

An algorithm for conducting hypotheses testing based on Neyman-Pearson Lemma. I. The case of continuous univariates

S. Parthasarathy and K. Sekar

Department of Crystallography and Biophysics, University of Madras, Madras 600 025, India

Neyman-Pearson Lemma provides a method for constructing the best critical region in hypotheses testing problems involving two simple hypotheses H_0 and H_1 . In many applications H_0 and H_1 are expressed as probability density functions $f_0(x)$ and $f_1(x)$. In practical situations the functions $f_0(x)$ and $f_1(x)$ are either too complicated to be amenable for straightforward theoretical treatment by Neyman-Pearson Lemma or given only as a table of numerical values. An algorithm to conduct the Neyman-Pearson test for such a complicated situation is proposed. A Fortran program, called NPTEST, for imple-

menting the algorithm has also been developed and tested.

In many hypotheses testing problems one has to decide between two probability density functions (PDF) $f_0(x)$ and $f_1(x)$ and the Neyman-Pearson Lemma¹ can be used to construct the best critical region for the test. In many practical situations the PDF $f_0(x)$ defining the null hypothesis H_0 and the PDF $f_1(x)$ defining the alternative hypothesis H_1 are either too complicated to be amenable for straightforward theoretical treatment by Neyman-Pearson Lemma or are given only as a table of numerical values. For example, in statistical tests for determining the space group symmetry of a crystal one meets with such a situation. It would therefore be quite useful to develop an algorithm for conducting the Neyman-Pearson test for such complicated situations and in this paper we shall describe one such algorithm. A Fortran program, called NPTEST, for implementing the algorithm has also been developed and tested using random samples generated from a few known probability distributions. NPTEST has also been tested in a few crystallographic situations related to space group determination. The details of these tests are also presented.

In the present paper we shall assume x to be a 1dimensional continuous random variable. For the convenience of numerical calculation we shall assume $f_0(x)$ and $f_1(x)$ to be defined in a finite interval a < x < b. If the PDF's under H_0 and H_1 are themselves defined on a finite interval, the choice of a and b are trivial. However if the PDF's under H_0 and H_1 are defined on an infinite interval, the user has to specify a suitable finite interval (a,b) for conducting the test. In this case it is advisable to choose a and b such that the probability mass of the original distributions included in that interval is as large as possible. If the PDF's under H_0 and H_1 are defined on an infinite interval, then NPTEST computes the truncated distributions applicable to the interval (a,b) and uses them in the calculation.

Consider the following hypotheses testing problem:

$$H_0: f(x) = f_0(x)$$

 $H_1: f(x) = f_1(x)$ (1)

Let $\{x_1, x_2, \dots, x_n\}$ be a random sample used in the test. We shall assume that, for any given value of x, the values of $f_0(x)$ and $f_1(x)$ can be calculated to a desired accuracy. In view of the Neyman-Pearson Lemma¹, the best critical region (BCR) is determined by

$$\prod_{j=1}^{n} \left[\frac{f_1(x_j)}{f_0(x_j)} \right] \ge k \,,$$
(2)

where k is a non-negative constant. Taking the natural logarithms of (2) and then dividing both sides by n, we obtain

$$(1/n) \sum_{k=1}^{n} t_{j} \ge (\ln k)/n$$
, (3)

where t is defined to be

$$t = \ln \left[\frac{f_1(x)}{f_0(x)} \right]. \tag{4}$$

Defining T and c by

$$T = \langle \ln \left[\frac{f_1(x)}{f_0(x)} \right] \rangle = \langle t \rangle, \ c = (\ln k)/n, \tag{5}$$

we can rewrite (3) compactly as

$$T \ge c$$
 (6)

T defined in (5) is the test statistic of the Neyman-Pearson test. Let $f(T|H_0)$ be the PDF of T when H_0 is true. The BCR of size α (i.e., the level of significance α) is given by the inequality

$$T \ge T_c$$
, (7)

where the value of T_c is to be chosen such that

$$P(T \ge T_c | H_0) = \alpha . \tag{8}$$

Thus the best test for the present situation is the one that is based on the statistic T with the region $T \ge T_c$ as the critical region. For a given α , the value of T_c satisfying (8) can be obtained by Monte Carlo Method as shown below.

To obtain the critical value T_{ε} for a given $\alpha (=0.05,$ say), we must first obtain the sampling distribution of Tunder the null hypothesis H_0 . This can be done numerically using the technique of inverse transform² by the Monte Carlo method and it involves the following steps: (i) Generating a value of the random variable x from the distribution given by the PDF $f_0(x)$ by the inverse transform method (see the next paragraph for more details); (ii) Using this value x compute the value of t(x) defined in (4); (iii) Repeating steps (i) and (ii) r times (r = 500, say) and obtaining r values of t (denoted by t_1, t_2, \dots, t_r). (iv) Computing the average value $\left(\sum_{i=1}^{r} t_i\right)/r$ and denoting it by T_1 . T_1 , thus obtained, is a sample value from $f(T|H_0)$; (v) Repeating the steps (i)-(iv) p times (p = 500, say) and obtaining the values T_1, T_2, \dots, T_p . These represent a random sample of size p from $f(T|H_0)$; (vi) Arranging the values T_1, T_2, \dots, T_p in decreasing order and obtaining the set of values T'_1, T'_2, \dots, T'_p , where $T'_1 \ge T'_2 \ge \dots \ge T'_p$;

Step (i) in the above procedure needs some explanation and we shall consider this aspect presently. Let $F_0(x)$ be the cumulative distribution function of x for the truncated PDF of x valid for the finite interval $a \le x \le b$. According to the principle of probability integral transformation³, the random variable u defined by

(vii) Since there are 500 values of T in all, it follows that

 $(T'_{25} + T'_{26})/2$ may be taken to be the solution of (8)

$$u = F_0(x) \tag{9}$$

is uniformly distributed in the interval $0 \le u \le 1$. For any given value u' of u the value of the inverse of the function in (12) can be obtained numerically by solving the equation

$$F_0(x) - u' = 0$$
 (10)

by the Newton-Raphson method* as

corresponding to $\alpha = 0.05$.

$$x_{m+1} = x_m - \frac{\{F_0(x_m) - u'\}}{f_0(x_m)}, m = 0, 1, 2 \cdots,$$
 (11)

where we have used the fact that $f_0(x)$ is the deviation of $F_0(x)$. The initial approximation x_0 needed for the Newton-Raphson algorithm in (11) may be obtained by the following steps: (i) Generating 501 equally spaced values of x in [a, b] and letting these be denoted by x_i , j=1 to 501. Computing the values of $F_0(x_j)$ (= u_j , say), j=1, 501. The ordered pairs (x_j, u_j) , j=1 to 501, provide a discrete representation of the function defined in (9). Evidently the ordered pairs (u_i, x_i) , j=1 to 501, provide a discrete representation of the inverse of the function defined in (12); (ii) For the ν (=43, say) numbers* $u_k' = 0$, 0.05, 0.1, ..., 0.9, 0.905, 0.910, ..., 0.98, 0.9825, 0.9850, ..., 1.0 computing the corresponding x_k' , k=1 to v, by using linear interpolation on the set $\{(u_j,$ x_i), j=1 to 501} obtained in step (i). The ordered pairs (u_k', x_k') , k=1 to v, thus obtained also provide a discrete representation of the inverse of the function defined in (9). This smaller set (i.e., $v \ll 501$) is more

convenient for obtaining the value of x for any given value of u by linear interpolation: (iii) Finding the interval in which the number u' lies and using it to find x_0 by linear interpolation on the set $\{(u_k', x_k'), k=1 \text{ to } v\}$.

A Fortran program, called NPTEST, has been written incorporating the above numerical calculations**. The program requires the following information as input: (i) A table of values of $f_0(x)$ and $f_1(x)$ for different values of x, (ii) The sample values $\{x_i\}$ for the test. It computes tables of values of the PDFs of x under H_0 and H_1 at 50 equally-spaced grid points in the interval (a,b). It then fits these two truncated PDFs with interpolating cubic splines. It calculates the cumulative function of x for the truncated distributions of x under H_0 and H_1 at the 50 grid points by numerical integration. It then computes the critical value of the test statistic T by Monte Carlo method. The program then conducts the test and prints out the unique decision arrived at.

The algorithm was tested for 3 hypothetical cases. In each case (see Table 1) a random sample of size 700 was generated from a specific distribution using the principle of random variable generation³. While the

^{**}The listing of NPTEST can be had from the authors on request.

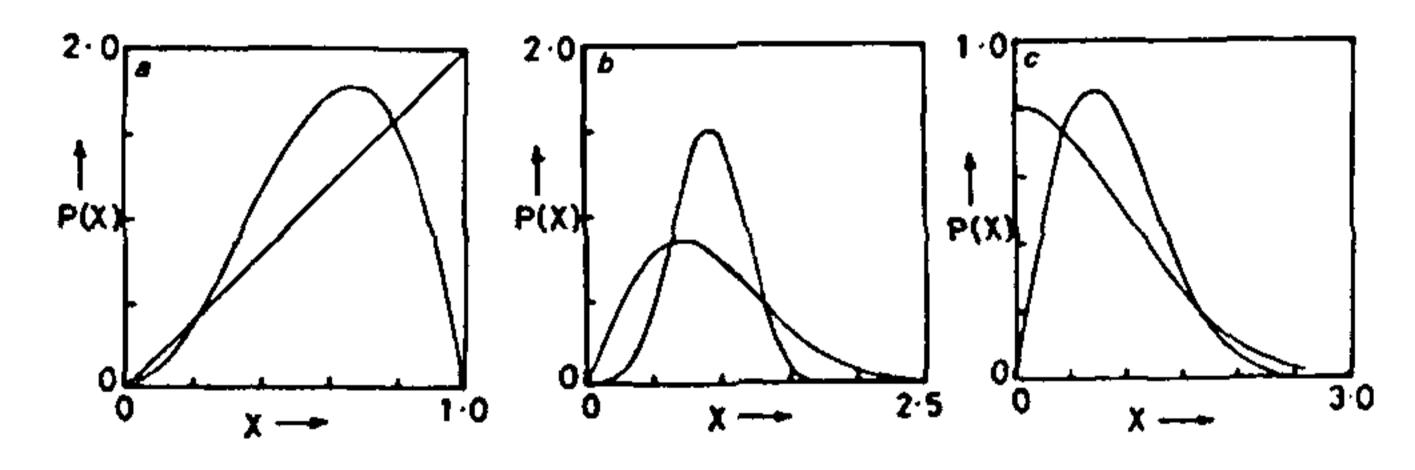


Figure 1. The graphs of the PDFs used to specify H_0 and H_1 for three cases. Curves (1 a), (1 h) and (1 c) are for the three cases considered in Table 1.

Table 1. Description of the three hypothetical cases used to test NPTEST

Case	PDFs used in hy testing	potheses	Domain of definition of the PDFs	Ref. fig.	The PDF used to generate the sample
	$f_{\alpha}(\mathbf{x})$	$f_1(x)$			
l 2	$\frac{2x}{2x \exp(-x^2)}$	$\frac{12x^{2}(1-x)}{4x^{3}\exp(-x^{4})}$	$0 \le x \le 1$	1 a 1 h	$f_{\phi}(x)$ $f_{\phi}(x)$
3	$\sqrt{2} \ln \exp(-x^2/2.0)$	$2x\exp(-x^2)$	$\chi \ge 0$	1.0	$f_1(x)$

Note: (i) For case 1, $f_0(x)$ is the PDI of the beta distribution's with parameters m-2 and n-1 and $f_1(x)$ is the PDI of the beta distribution with parameters m-3 and n-2 (ii) For case 2, $f_0(x)$ is the PDI of the Weibull distribution's with parameters n-2 and n-1 and $f_1(x)$ is the PDI of the Weibull distribution with parameters n-4 and n-1 (iii) I or case 3, $f_0(x)$ and $f_1(x)$ are the PDIs of the normalized structure factor magnitude for the centric and accentic Wilson distributions'' which occur in the field of x ray crystallography.

^{*}The grid-point-spacings, in the u'-space are taken to be 0.05, 0.005, 0.0025 for $0 \le u' \le 0.9$, $0.9 \le u' \le 0.98$ and $0.98 \le u' \le 1$ respectively. Such a choice has been made to obtain sufficiently accurate values for x_0 in the numerical process.

Table 2	Results obtained	from NPTEST	for the 3 c	ases described in	Table 1
---------	------------------	-------------	-------------	-------------------	---------

Case	Critical region	Observed value of the test statistic	Decision obtained regarding the PDF of the sample	PDF used to generate the sample
1	[-0149	0 126	$f_0(x)$	$f_{0}(x)$
2	f - 0 587	0.583	$f_0(x)$	$f_{\theta}(x)$
3	Č-0 148	-0.225	$f_1(x)$	$f_1(x)$

Note: $\alpha = 0.05$. $\{x_0\}$ stands for the interval $0 \le x \le x_0$.

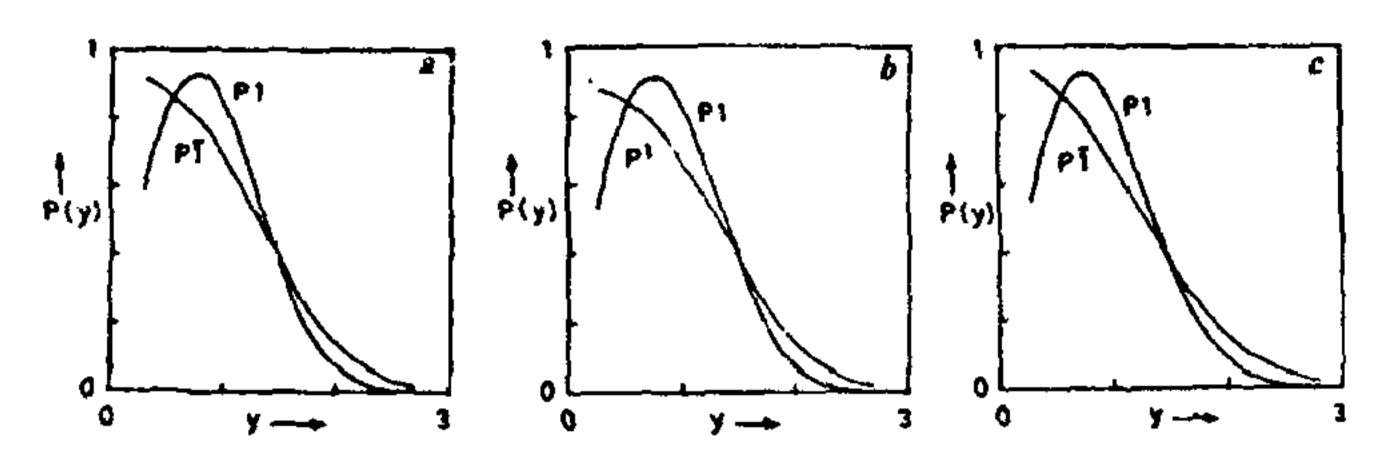


Figure 2. The graphs of the PDFs of y for three crystallographic situations. Curves (2a), (2b) and (2c) are for the crystals 1, 2 and 3 of Table 3.

PDFs $f_0(x)$ and $f_1(x)$ for the first case are defined on the finite interval $0 \le x \le 1$, those for the second and the third are defined on the interval $0 \le x < \infty$. For the latter two cases the theoretical truncated distributions valid for the interval $0 \le x < b$ were used in the test. The values of b for the second and third cases were determined such that the probability $Pr(0 \le x \le b)$, calculated using that PDF which tends to zero less slowly, is greater than or equal to 0.999. The sample data which lies in the interval $0 \le x \le b$ was used in the test for the cases 2 and 3. The results obtained from the tests conducted using NPTEST for these three cases are given in Table 2 which shows that correct decisions are obtained in all the cases.

The algorithm was also tested for three situations met with in X-ray crystallography. The PDF of the normalized structure factor magnitude y for a triclinic crystal containing 2 heavy atoms per unit cell besides a large number of light atoms is given by⁷

$$P(y) = (1/\sigma_2) (2/\pi)^{3/2} \exp(-y_N^2/2\sigma_2^2)$$

$$\left[\int_0^{\pi/2} \exp[-\sigma_1^2 \sin^2 \psi/\sigma_2^2] \times \cos h \left[\sqrt{2} \sigma_1 y_N \sin \psi/\sigma_2^2 \right] d\psi \right], \qquad (12)$$

if the space group of the crystal is P I and by

$$P(y) = (4y_N/\pi\sigma_2^2) \exp(-y_N^2/\sigma_2^2)$$

$$\left[\int_0^{\pi/2} \exp[-(2\sigma_1^2 \sin^2 \psi/\sigma_2^2) \times \right]$$

$$I_0 [2\sqrt{2}\sigma_1 y_N \sin \psi/\sigma_2^2] d\psi$$
, (13)

if the space group of the crystal is P1. The parameter σ_i^2 is a measure of the contribution of the heavy atoms to the local average intensity relative to that from all the atoms in the unit cell of the crystal and $\sigma_2^2 = 1 - \sigma_1^2$. A crystallographer meets with the problem of deciding whether the space group of the crystal under investigation is P1 or $P\overline{1}$ from the normalized structure amplitude data (called, y-values) which could be obtained from the intensities of X-ray reflections. This problem can be solved by testing whether the observed y-data fits better either (12) or (13). Since σ_1^2 is a slowly varying function of the (sin θ/λ) where 2θ is the angle of scattering and λ is the wavelength of X-rays, an average value of σ_1^2 is to be determined for the reflections used in the test and this average value is to be used to compute the table of values of the PDFs of y corresponding to the null and alternative hypotheses. The tests were conducted by using the truncated PDFs, the limits a and b being determined from an analysis of y-values. Relevant details of the three crystals used in the tests are given in Table 3 and the theoretical PDFs of y corresponding to the hypothesis H_0 (i.e., space

Table 3. Details of the three crystals⁸⁻¹⁰ used to test NPTEST

Crystal	Asymmetric unit	Space group	a	ь	σ_1^2
1	C, H ₁₂ NO ₃ Cl	ΡĪ	0.314	2,700	0.53
2	$C_{39}H_{52}N_3O_{12}Br$	PŢ	0.286	2,700	0.56
3	C ₁₆ H ₂₂ NOCl	PĪ	0.287	2.800	0.44

Table 4.	Results obtained from	NPTEST for the three	crystals described in Table 3
----------	-----------------------	----------------------	-------------------------------

Crystal	Observed value Critical of the test region statistic		Decision obtained regarding the space group of the crystal	Actual space group of the crystal
1	[-0.020	0.038	ΡĪ	ΡĪ
2	(-0.021)	0.032	PΪ	ΡĪ
3	[-0.023]	0.047	ΡĪ	PŢ

Note: $\alpha = 0.05$. '[x_0 ' stands for the interval $0 \le x \le x_0$.

group of the crystal is P1) and H_1 (i.e., space group of the crystal is P1) are shown in Figure 2. The results obtained from the tests are given in Table 4. It is seen that correct results are obtained in all the three cases.

- 1. Hoel, P. G., Port, S. C. and Stone, C. J., Introduction to Statistical Theory, Houghton Mifflin Co., Boston, 1971, pp. 56-60.
- 2. Rubinstein, R. Y., Simulation and the Monte Carlo Method, John Wiley and Sons, New York, 1981.
- 3. Sobol, I. M., The Monte Carlo Method, Mir Publishers, Moscow, 1975, pp. 32-34.
- 4. Ralson, A. and Rabinowitz, P., A First Course in Numerical Analysis, Mc Graw Hill, Intl. Book Co. Tokyo, 1981, p. 346.
- 5. Eadie, W. T., Drijard, D., James, F. E., Roos, M. and Sadoulet, B.,

- Statistical Methods in Experimental Physics, North-Holland Publishing Co., Amsterdam, 1971, p. 73 & p. 83.
- 6. Wilson, A. J. C., Acta Crystallogr., 1949, 2, 318-321.
- 7. Srinivasan, R. and Parthasarathy, S., Some Statistical Applications in X-ray Crystallography, Pergamon Press, New York, 1976, pp. 38-42.
- 8. Brehm, L. and Honore, T., Acta Crystallogr., 1978, B34, 2359-2361,
- 9. Sheldrick, W. S., Brokenstein, A. and Engel, J., Acta Crystallogr., 1978, B34, 2055-2058.
- 10. Herbert, H., Acta Crystallogr., 1978, B34, 3128-3131.

ACKNOWLEDGEMENT. Thanks are due to the Council of Scientific and Industrial Research, India for the award of Senior Research Fellowship to K. S.

Received 26 March 1992; revised accepted 22 July 1992

Self-diffusion coefficient of 86-Rb in mica-rich soils

B. B. Mishra*, S. K. Ghosh, P. Sachdev, M. S. Sachdev, D. L. Deb and R. P. Arora

Nuclear Research Laboratory, Indian Agricultural Research Institute, New Delhi 110 012, India

*Present address: Soil Survey and Land Use Planning Scheme, Rajendra Agricultural University, Sabour Campus, Sabour 813 210, India

The self-diffusion coefficient of 86-Rb in four surface soils derived from mica-rich parent materials was determined at field capacity. Its values ranged from 0.54×10^{-10} to 1.674×10^{-10} cm² sec⁻¹ at zero level K and from 1.788×10^{-10} to 7.02×10^{-10} cm² sec⁻¹ at 50 ppm level K. These values were found much lower than the values reported elsewhere earlier. Enrichment of these soils with 50 ppm carrier-K resulted in a tremendous increase in the Da-Rb values.

The movement of K through diffusion is important in soils¹. Potassium does not have any suitable isotope to use in labelling soil. The ⁴²K and ⁴³K have half-lives only in hours, while enrichment of natural isotope, ⁴⁰K and the stable isotope ⁴¹K is very expensive. The rubidium was tested as a tracer of K in soil studies². The process of Rb uptake from soil by plant roots is similar to that of K uptake^{3,4}. The self-diffusion of ⁸⁶Rb was found linearly correlated positively with soil

moisture⁵. In view of this, an experiment was conducted to study the self-diffusion coefficient of ⁸⁶Rb in soils derived from mica-rich parent materials. Such information particularly for the soils of India is meagre.

Surface soils of four representative pedons of Giridih and Munger districts of Bihar were collected. Half-cell technique as outlined by Sen and Deb⁶ was used to study the Rb-diffusion, at two K-levels i.e. 0 and 50 ppm K. The tagging of soils with 86-Rb was performed taking 1 µCi/g of soil. A calculated amount of water corresponding to water content at 0.3 bar (field capacity) was added. Activity of tagged and untagged soil portions was measured using a G. M. counter. All the determinations were made in duplicate. The apparent self-diffusion coefficient of ⁸⁶Rb was then calculated by using the integrated form of the equation derived by Phillips and Brown⁷ and presented in Table

The self-diffusion coefficient ⁸⁶Rb in all four soils at zero level K ranged from 0.54×10^{-10} to 1.674×10^{-10} cm² s⁻¹. The enrichment of these soils with 50 ppm carrier-K before determining the Da Rb, however, resulted in a tremendous increase in the Da Rb values at field capacity. This increase in Da Rb appears to be possibly due to an increase in soil solution concentration of K. The values for Da Rb in these mica-rich soils were observed to be generally lower than the values reported earlier^{7,8}. Such low values of Da Rb might have been due to the fact that the soil clays were rich in