(2)

Talent 1987, 0312

55619

Statistical Methods for Environmental Pollution Monitoring

Richard O. Gilbert

Pacific Northwest Laboratory

APR 2 3 1997
em

VAN NOSTRAND REINHOLD
New York

10451

Dedicated to my parents, Mary Margaret and Donald I. Gilbert

Copyright © 1987 by **Van Nostrand Reinhold**
Library of Congress Catalog Card Number: 86-26758
ISBN 0-442-23050-8

Work supported by the U.S. Department of Energy under Contract No. DE-AC06-76RLO1830.

All rights reserved. No part of this work covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without written permission of the publisher.

Printed in the United States of America

Van Nostrand Reinhold
115 Fifth Avenue
New York, New York 10003

Chapman & Hall
2-6 Boundary Row
London SE1 8HN, England

Thomas Nelson Australia
102 Dodds Street
South Melbourne, Victoria 3205, Australia

Nelson Canada
1120 Birchmount Road
Scarborough, Ontario M1K 5G4, Canada

16 15 14 13 12 11 10 9 8 7 6 5 4

Library of Congress Cataloging-in-Publication Data
Gilbert, Richard O.
Statistical methods for environmental pollution monitoring.
Bibliography: p.
Includes index.
1. Pollution—Environmental aspects—Statistical methods. 2. Environmental monitoring—Statistical methods. I. Title.
TD193.G55 1987 628.5'028 86-26758
ISBN 0-442-23050-8

10 Locating Hot Spots

Chapters 4 through 9 have discussed sampling designs for estimating average concentrations or total amounts of pollutants in environmental media. Suppose, however, that the objective of sampling is not to estimate an average but to determine whether "hot spots," or highly contaminated local areas are present. For example, it may be known or suspected that hazardous chemical wastes have been buried in a land fill but its exact location is unknown. This chapter provides methods for answering the following questions when a square, rectangular, or triangular systematic sampling grid is used in an attempt to find hot spots:

What grid spacing is needed to hit a hot spot with specified confidence?
For a given grid spacing, what is the probability of hitting a hot spot of specified size?
What is the probability that a hot spot exists when no hot spots were found by sampling on a grid?

This discussion is based on an approach developed by Singer (1972, 1975) for locating geologic deposits by sampling on a square, rectangular, or triangular grid. He developed a computer program (ELIPGRID) that was used by Zirschky and Gilbert (1984) to develop nomographs for answering the preceding three questions. These nomographs are given in Figures 10.3, 10.4, and 10.5. We concentrate here on single hot spots. Some approaches for finding multiple hot spots are discussed by Gilbert (1982) and Holoway et al. (1981).

The methods in this chapter require the following assumptions:

1. The target (hot spot) is circular or elliptical. For subsurface targets this applies to the projection of the target to the surface (Fig. 10.1).
2. Samples or measurements are taken on a square, rectangular, or triangular grid (Fig. 10.2).
3. The distance between grid points is much larger than the area sampled, measured, or cored at grid points—that is, a very small proportion of the area being studied can actually be measured.
4. The definition of "hot spot" is clear and unambiguous. This definition implies that the types of measurement and the levels of contamination that constitute a hot spot are clearly defined.

NOV 02 '98 09:35AM MK ENVIRONMENTAL P.4

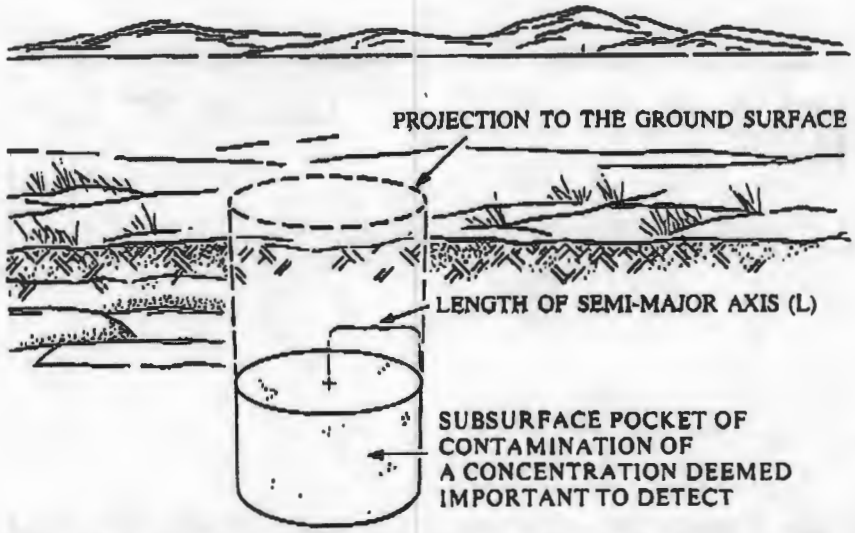

Figure 10.1 Hypothetical subsurface pocket of contamination (after Gilbert, 1982, Fig. 1).

5. There are no measurement misclassification errors—that is, no errors are made in deciding when a hot spot has been hit.

Parkhurst (1984) compared triangular and square grids when the objective is to obtain an unbiased estimate of the density of waste clusters in a hazardous waste site. He showed that the triangular grid was more likely to provide more

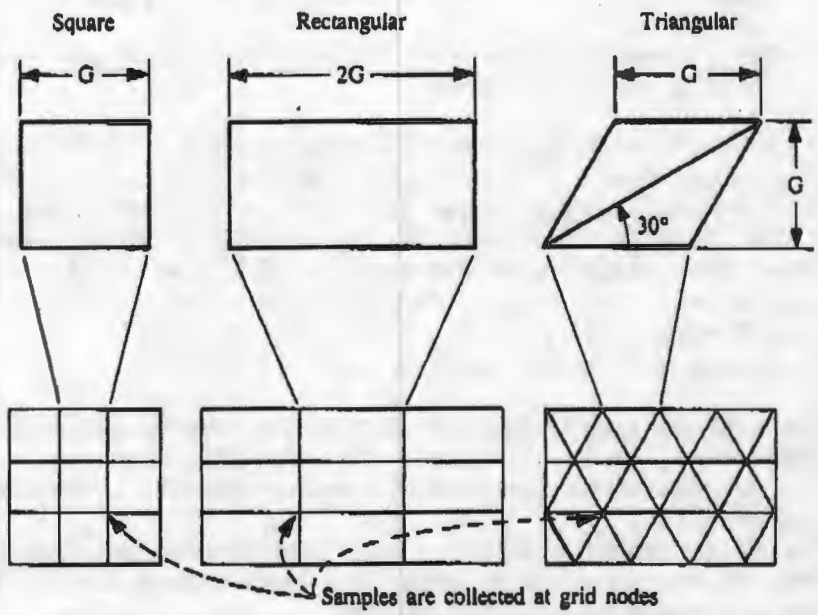

Figure 10.2 Grid configurations for finding hot spots (after Zirschky and Gilbert, 1984, Fig. 1).

NOV 02 '98 09:36AM MK ENVIRONMENTAL P.5

information than the square grid. He also concluded that if the waste clusters are expected to follow an unknown but regular pattern, the wells should be drilled at randomly selected locations. But for randomly located clusters, a triangular or square grid is preferred.

10.1 DETERMINING GRID SPACING

The grid spacing required to find a hot spot of prespecified size and shape with specified confidence may be determined from the following procedure:

1. Specify L, the length of the semimajor axis of the smallest hot spot important to detect (see Fig. 10.1). L is one half the length of the long axis of the ellipse.
2. Specify the expected shape (S) of the elliptical target, where

$$S = \frac{\text{length of short axis of the ellipse}}{\text{length of long axis of the ellipse}}$$

   Note that $0 < S \leq 1$ and that $S = 1$ for a circle. If S is not known in advance, a conservative approach is to assume a rather skinny elliptical shape, perhaps $S = 0.5$, to give a smaller spacing between grid points than if a circular or "fatter" ellipse is assumed. That is, we sample on a finer grid to compensate for lack of knowledge about the target shape.
3. Specify an acceptable probability (β) of not finding the hot spot. The value β is known as the "consumer's risk." To illustrate, we may be willing to accept a $100\beta\% = 20\%$ chance of not finding a small hot spot, say one for which $L = 5$ cm. But if L is much larger, say $L = 5$ m, a probability of only $\beta = 0.01$ (1 chance in 100) may be required.
4. Turn to Figures 10.3, 10.4, or 10.5 for a square, rectangular, or triangular grid, respectively. These nomographs give the relationship between β and the ratio L/G, where G is the spacing between grid lines (Fig. 10.2). Using the curve corresponding to the shape (S) of interest, find L/G on the horizontal axis that corresponds to the prespecified β. Then solve L/G for G, the required grid spacing. The total number of grid points (sampling locations) can then be found because the dimensions of the land area to be sampled are known.

For elliptical targets ($S < 1$) the curves in Figures 10.3, 10.4, and 10.5 are average curves over all possible orientations of the target relative to the grid. Singer (1975, Fig. 1) illustrates how the orientation affects the probability of not hitting the target. If the orientation is known, Singer's (1972) program will give the curves for that specific orientation.

EXAMPLE 10.1

Suppose a square grid is used and we want to take no more than a $100\beta\% = 10\%$ chance of not hitting a circular target of radius $L = 100$ cm or larger. Using the curve in Figure 10.3 for $S = 1$, we find $L/G = 0.56$ corresponds to $\beta = 0.10$. Solving for G yields $G = L/0.56 = 100$ cm$/0.56 \cong 180$ cm. Hence, if cores are taken

NOV 02 '98 09:36AM MK ENVIRONMENTAL P.6

NOV 02 '98 09:37AM MK ENVIRONMENTAL P.7

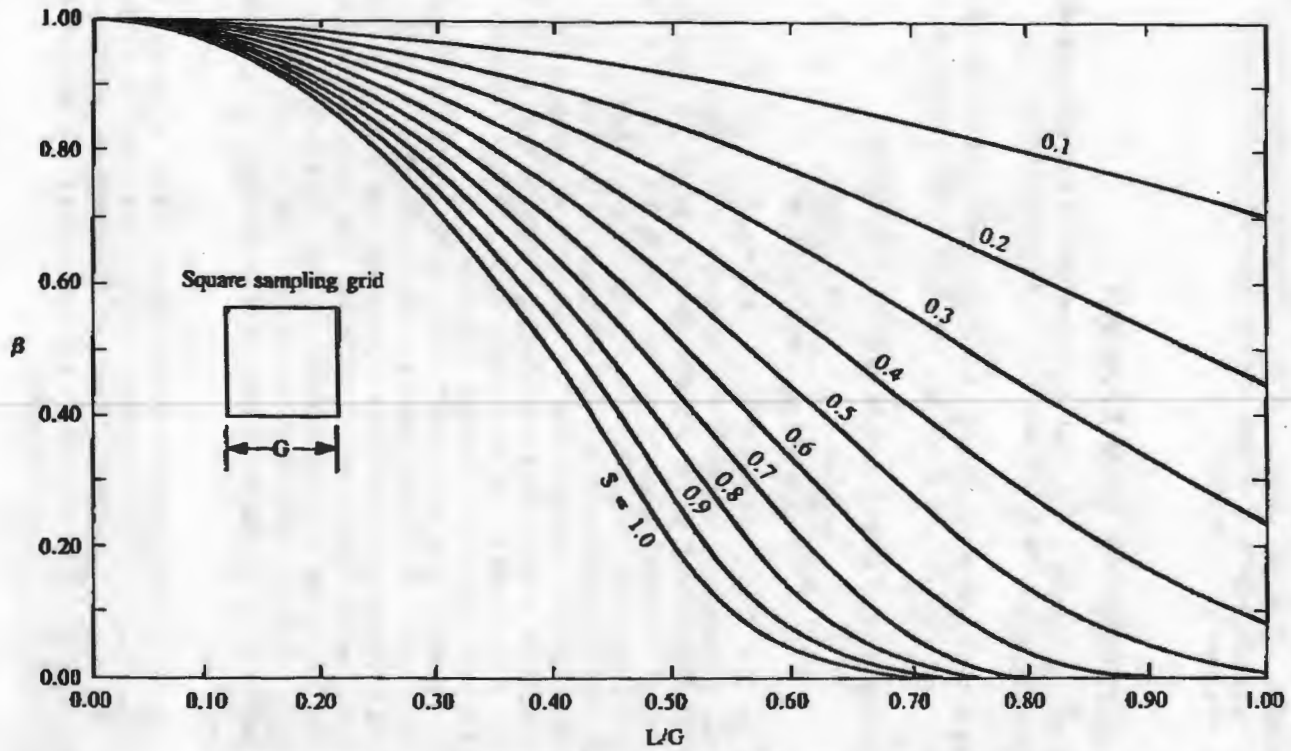

Figure 10.3 Curves relating L/G to consumer's risk, β, for different target shapes when sampling is on a square grid pattern (after Zirschky and Gilbert, 1984, Fig. 3).

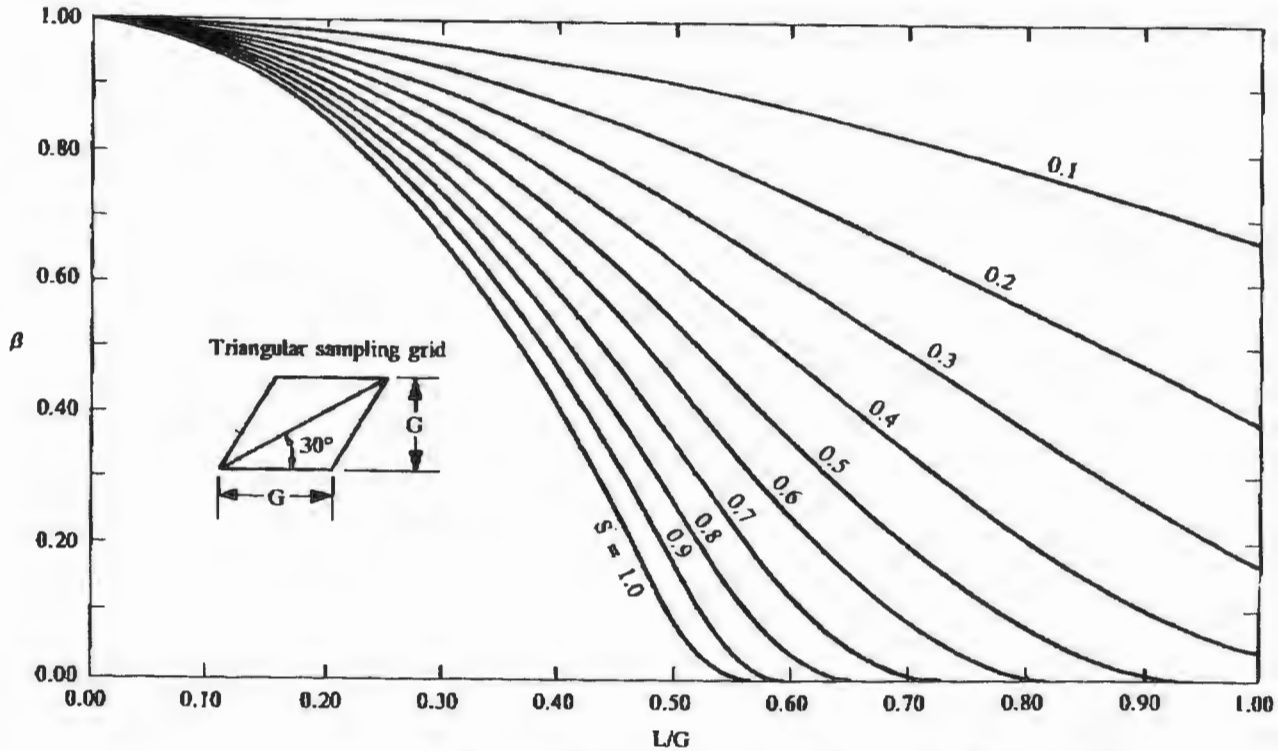

Figure 10.5 Curves relating L/G to consumer's risk, β, for different target shapes when sampling is on a triangular grid pattern (after Zirschky and Gilbert, 1984, Fig. 4).

NOV 02 '98 09:37AM MK ENVIRONMENTAL

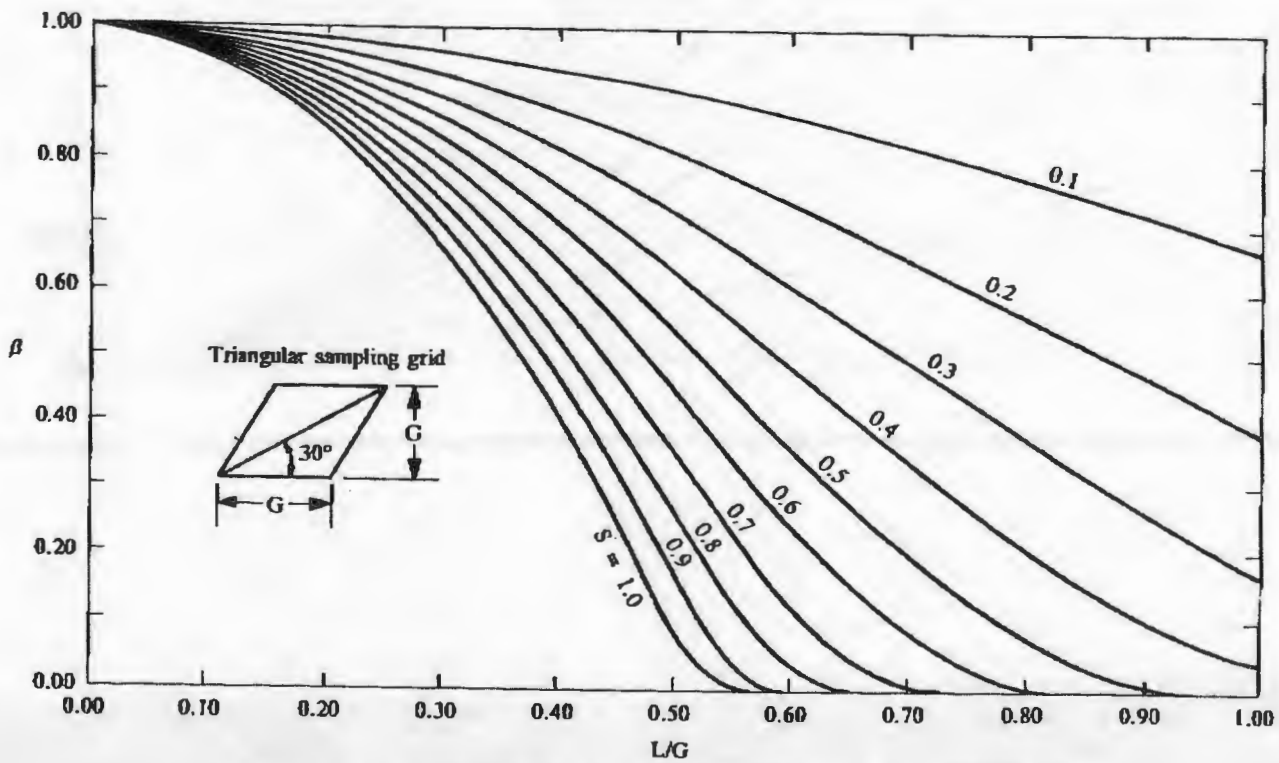

Figure 10.5 Curves relating L/G to consumer's risk, β, for different target shapes when sampling is on a triangular grid pattern (after Zirschky and Gilbert, 1984, Fig. 4).

NOV 02 '98 09:37AM MK ENVIRONMENTAL

on a square grid with spacing of 180 cm, we are assured the probability is only 0.10 (1 chance in 10) of not hitting a circular target that is 100 or more cm in radius.

10.2 SIZE OF HOT SPOT LIKELY TO BE HIT

Figures 10.2, 10.3, and 10.4 can also be used to find the maximum size hot spot that can be located for a given cost and consumer's risk. Suppose, for example, we can afford to take measurements at no more than 25 locations on a square grid system. What size elliptical target (characterized by L) can we expect to find with confidence $1 - \beta$ (the probability of hitting a target at least once)? The general procedure is to specify β, G, and S, then use the curves to solve for L.

EXAMPLE 10.2

Suppose our budget allows taking measurements at $n = 25$ locations on a square grid pattern. Suppose also that a grid spacing of $G = 200$ cm covers the area of interest. What size circular target can we be at least 90% confident of detecting—that is, for which the probability of not hitting the target is $\beta = 0.10$ or less? Using $S = 1$ in Figure 10.3, we find $L/G = 0.56$ for $\beta = 0.10$. Hence, $L = (200 \text{ cm})(0.56) = 112$ cm. Therefore, we estimate that a circle with a radius of 112 cm or larger has no more than a 10% chance of not being hit when using a square grid spacing of 200 cm. If the circular target has a radius L less than 112 cm, the probability of not locating it will exceed 0.10. Conversely, if $L > 112$ cm, the probability of not locating it will be less than 0.10. If we require only a 50% chance of hitting the target (i.e., $\beta = 0.50$), the curve for $S = 1$ gives $L/G = 0.4$ or $L = (200 \text{ cm})(0.4) = 80$ cm.

By computing L as in Example 10.2 for different values of β and G, we can generate curves that give the probability of hitting a circular or elliptical target of any size. These curves for grid spacings of 100, 200, and 300 distance units for two target shapes, $S = 1$ and 0.5, are given in Figure 10.6.

For example, suppose the target is circular ($S = 1$) and the grid spacing is $G = 100$ units. Then the probability β that we do not hit a circular target of radius $L = 50$ units (same units as G) is about 0.2. If the target is smaller, say $L = 20$ units, then β is larger, about 0.87.

10.3 PROBABILITY OF NOT HITTING A HOT SPOT

Figures 10.3–10.5 can also be used to estimate the consumer's risk β of not hitting a hot spot of given size and shape when using a specified grid size.

NOV 02 '98 09:38AM MK ENVIRONMENTAL P.10

NOV 02 '98 09:38AM MK ENVIRONMENTAL P.11

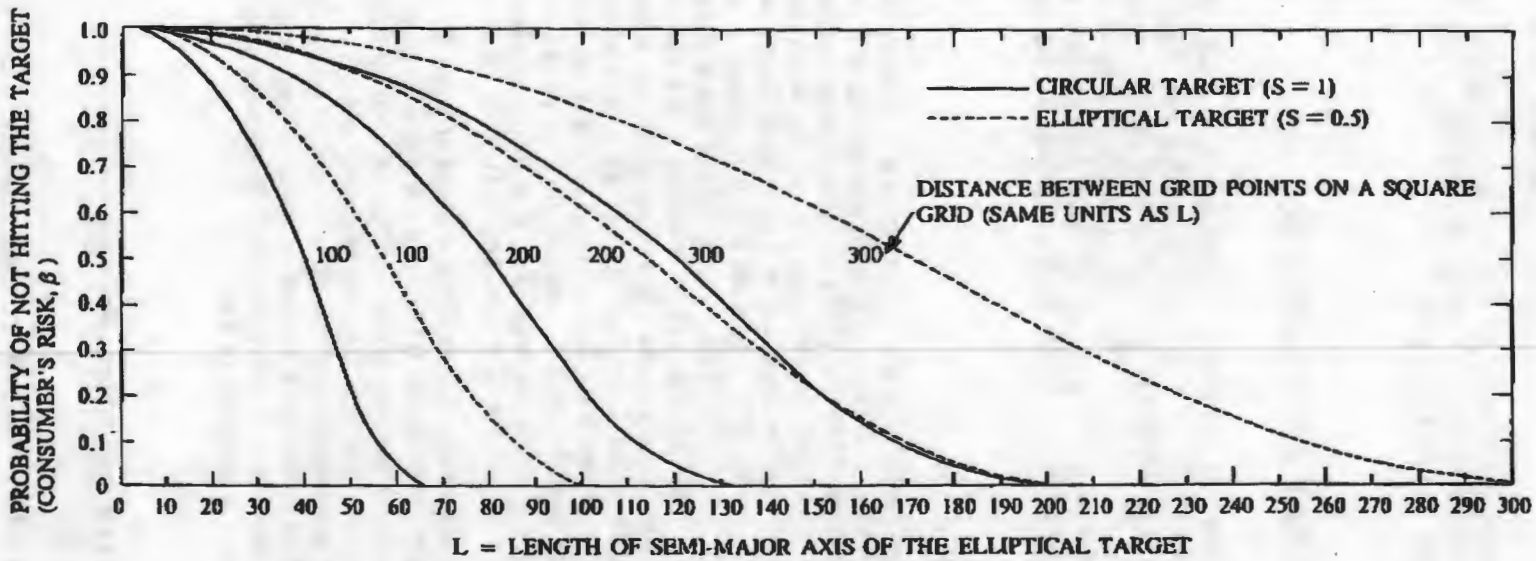

Figure 10.6 Curves relating L to the probability (β) of not hitting the circular or elliptical target. Curves are for a square grid spacing (after Gilbert, 1982, Fig. 4).

EXAMPLE 10.3

What is the average probability of not finding an elliptical hot spot that is twice as long as it is wide and for which the semimajor axis (L) is 40% as long as the spacing G between grid points? Suppose a rectangular sampling grid is used. Using $S = 0.5$ and $L/G = 0.40$ in Figure 10.4, we find β to be 0.87. Hence, there is about an 87% chance that this size and shape target would not be found by sampling at the grid points. The actual β could be somewhat smaller or larger than 0.87, depending on the orientation of the target relative to the grid.

10.4 TAKING PRIOR INFORMATION INTO ACCOUNT

Thus far we have assumed that a hot spot does actually exist. In practice, no such assurance may be warranted. Now we consider how prior information about the probability that a hot spot exists can be used to obtain a more realistic estimate of β. Let

A = event that a hot spot of size L or larger exists

B = event that a hot spot of size L or larger is hit by taking measurements on a grid.

Then the law of conditional probabilities (see, e.g., Fisz, 1963, p. 20) says that

$$P(B|A) = \frac{P(A, B)}{P(A)}$$

= probability that a hot spot of size L or larger is hit, *given* such a hot spot exists **10.1**

where

$P(A, B)$ = probability that a hot spot of size L or larger exists *and* is discovered by sampling on a grid

and

$P(A)$ = probability that a hot spot of size L or larger exists

Whenever there is doubt whether a hot spot of size L or larger exists, then $P(A, B)$ is of interest. From Eq. 10.1 we have

$$P(A, B) = P(B|A)P(A) \qquad \textbf{10.2}$$

Now $P(B|A)$ is just $1 - \beta$. Hence, $P(A, B)$ can be estimated by using Figures 10.3–10.5 and by specifying a value for $P(A)$. In many situations a hot spot of size L or larger will be known to exist so that $P(A) = 1$ and $P(A, B) = P(B|A)$. Then if a square grid is used, Figure 10.3 gives the final result. In other

NOV 02 '98 09:39AM MK ENVIRONMENTAL P.12