A COMPARATIVE STUDY OF SIMULATION TECHNIQUES FOR TWO DIMENSIONAL DATA HONOURING SPECIFIED EXPONENTIAL SEMIVARIOGRAMS

P. I. BROOKER[1] and M. A. STEWART[2]

Abstract

The effectiveness of four techniques for producing wide sense stationary data with exponential semivariograms is examined. Comparison is made primarily on the basis of the observed semivariograms. The LU decomposition of the covariance matrix appears to most accurately model specified semivariograms, whilst the more computationally efficient Matrix Polynomial approximation and Turning Bands methods may be more useful in practice.

1. Introduction

The production of spatially correlated data is an important tool for many simulation applications. Agriculture, forestry, mining and petroleum geology all have profited from being able to model the distribution of natural resources, using random function theory. This paper compares four techniques for producing wide sense stationary data with specified semivariograms. Of principal interest is the reliability of the technique, expense and complexity being a secondary matter.

The semivariogram γ, is an indicator of the correlation structure of a random function $Y(\vec{x})$, and is defined by

$$\gamma(\vec{h}) \stackrel{\text{def}}{=} \frac{1}{2} E\left[\left\{Y(\vec{x} + \vec{h}) - Y(\vec{x})\right\}^2\right],$$

where $\vec{x}$ and $\vec{x} + \vec{h}$ are two points in space separated by a displacement $\vec{h}$, and $Y(\vec{x})$ is a real valued function, in the case here the data generated, evaluated at the point $\vec{x}$. The dependence of γ upon the position $\vec{x}$ is deliberately omitted in

[1]Department of Geology and Geophysics, University of Adelaide, South Australia 5005.
[2]Teletraffic Research Centre, University of Adelaide, South Australia 5005.

this definition to emphasise the commonly used assumption of stationarity. This assumption requires that the probability distribution is translationly invariant. The validity of the assumption of stationarity warrants careful attention; however it does provide a useful working hypothesis and is often acceptable, if not on the full field then upon subregions from which useful conclusions may be drawn.

The simulation of a random function Y which is consistent with a specified semivariogram γ, is a computationally expensive process and a range of techniques exist for generating data which honours γ. This paper considers four techniques for generating such data,

(i)   The method of Turning Bands in Three Dimensions (Journel and Huijbregts [5]), TB3D. Only the case with 15 bands ([5]) is considered;

(ii)   The method Turning Bands in Two Dimensions (Brooker [1]), TB2D, using both 15 and 50 bands;

(iii)   The Cholelsky, or LU decomposition of the Covariance Matrix (Davis [4]), LU;

(iv)   The Matrix Polynomial approximation of the square root of the Covariance Matrix (Davis [3]), MPOLY, using both 8 and 15 coefficients (Brooker, Cock, and Stewart [2]);

and makes comparison of the degree to which they honour an exponential semivariogram on two dimensional data sets. It may be seen as an extension of [2], where the same four techniques were studied in relation to specified spherical semivariograms. The reader may refer to Mantoglou, and Wilson [6], and Tompson, Ababou, and Gelhar [7], for other studies involving the Turning Bands methods.

2. Exponential model

The exponential semivariogram is :

$$\gamma(h) = C_0 + C \left\{ 1 - \exp\left(\frac{-h}{a}\right) \right\},$$

where a is a constant, being one third of the effective range ([5] page 508) of the semivariogram, h is the length of the displacement $\vec{h}$, $C_0 + C$ is the sill, and C_0 represents the short range variability known as the nugget effect. Since this nugget may be added after the correlated component of the data is produced, C_0 is set to 0. C is also set to 1, for all calculations preformed here, since the resulting data can readily be scaled to any other value of C.
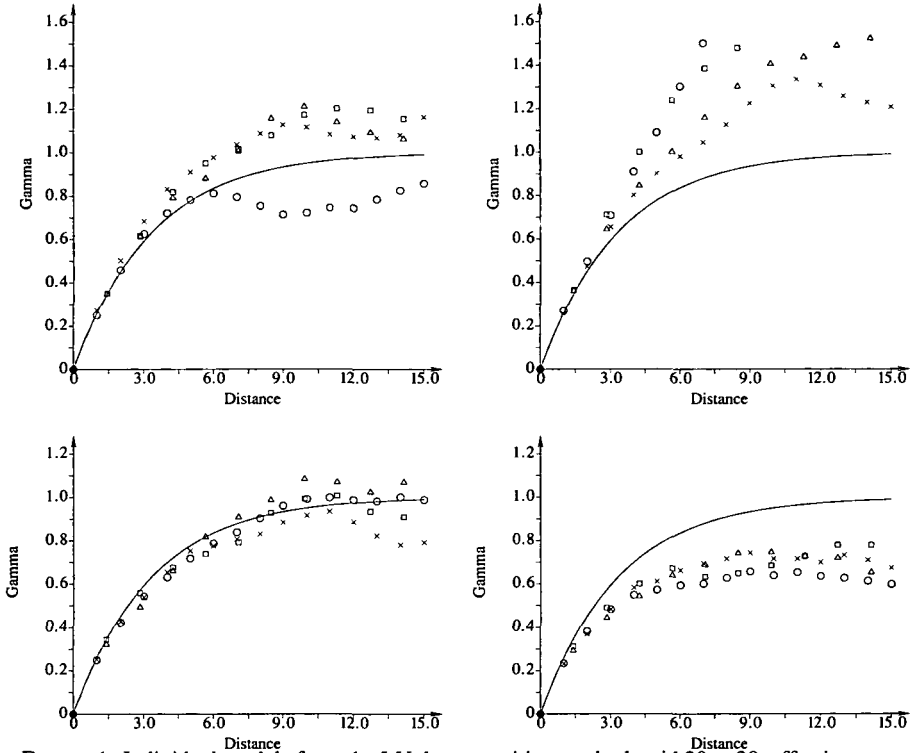
FIGURE 1. Individual models from the LU decomposition method, grid 30×30, effective range 10.

3. Results

This study investigates the effectiveness of each of the four techniques in generating data sets honouring specified exponential semivariograms with a number of effective ranges. The basis for comparison will be how reliably data generated on an $L \times L$ square grid, by each of the four techniques, honours the specified semivariogram. To perform this comparison an ensemble of 50 grids is generated for each effective range, and grid size. For each data set the standard estimator

$$\gamma^*(\vec{h}) = \frac{1}{2N(\vec{h})} \sum \left\{ Y(\vec{x} + \vec{h}) - Y(\vec{x}) \right\}^2,$$

of the semivariogram, is calculated in the N-S, E-W, NE-SW, and NW-SE directions. Figures 1 and 3 show four individual semivariograms for each of LU and TB3D, 30×30 grids with effective range 10, and include "ideal" and extreme examples.

From the resulting collection of experimental semivariograms the following calculations are made:
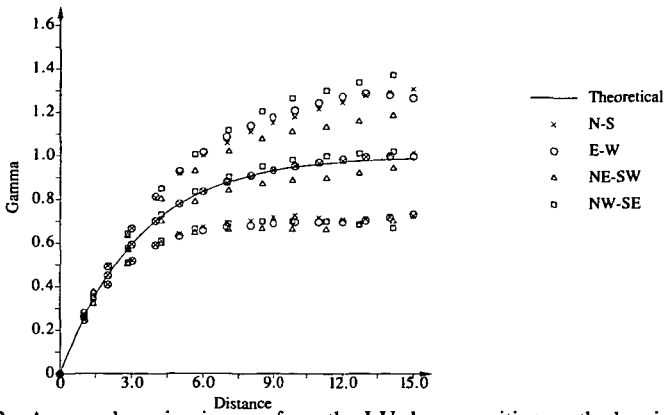
FIGURE 2. Averaged semivariogram from the LU decomposition method, grid 30 × 30, effective range 10.
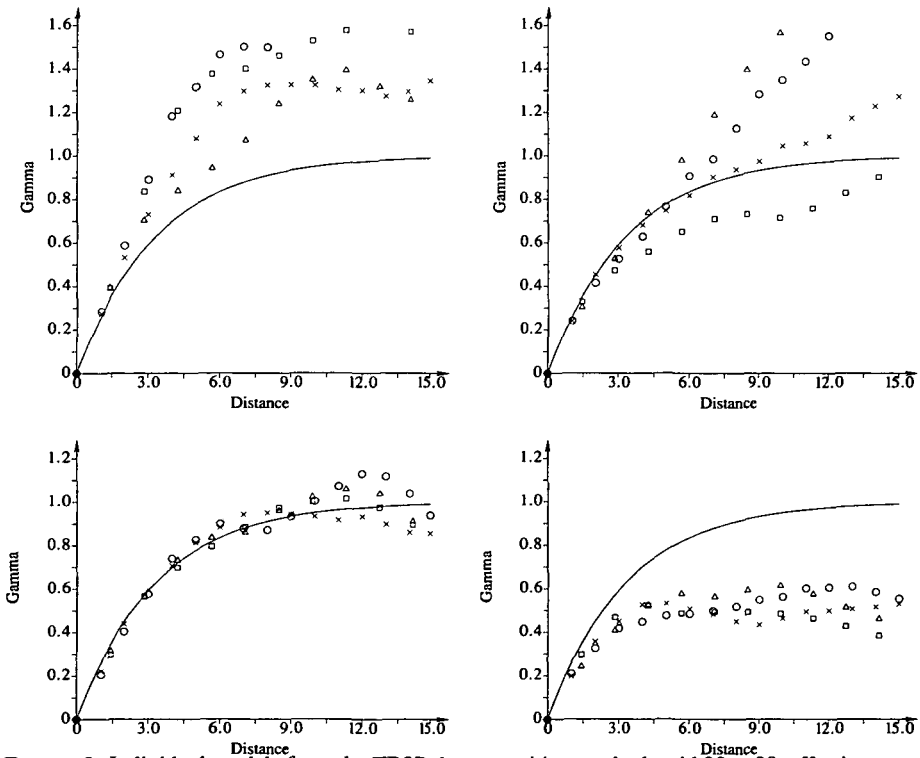

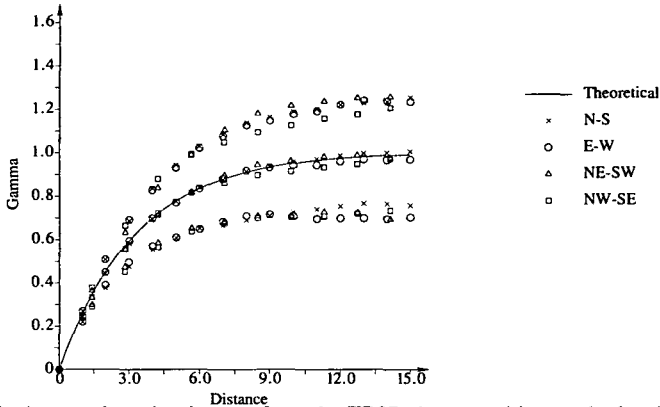FIGURE 3. Individual models from the TB3D decomposition method, grid 30 × 30, effective range 10.

FIGURE 4. Averaged semivariogram from the TB3D decomposition method, grid 30 × 30, effective range 10.

(i)   For each technique an averaged semivariogram $\overline{\gamma^*}$ is computed, by averaging the values of the standard estimator of the semivariogram γ^* at each lag $\vec{h}$; over the 50 grids (of specified range, and grid size).

(ii) Three measures of the degree to which the averaged semivariogram $\overline{\gamma^*}$ deviates from the specified semivariogram γ are computed, by calculating the average of the absolute difference between the averaged semivariogram $\overline{\gamma^*}$ and the specified semivariogram γ, in each of the four directions over the ranges $h \in [1, 3a]$, $h \in [1, L/2]$, and $h \in (3a, L/2]$. $L/2$ is the recommended maximum lag ([5], page 194) if experimental semivariograms are to be reliable.

(iii) At each of the lag distances h the standard deviation of the 50 values of the standard estimators γ^* corresponding to grids of constant range, grid size and method is computed.

(iv) Three measures of the variability in the experimental semivariograms γ^* are computed by averaging the standard deviations computed above (iii) over the ranges $h \in [1, 3a]$, $h \in [1, L/2]$ and $h \in (3a, L/2]$ for all grid sizes, effective ranges and methods.

Figures 2 and 4 depict the averaged semivariogram $\overline{\gamma^*}$, for the full ensembles, four of whose individual semivariograms are seen in Figures 1 and 3 respectively. The average plus or minus one standard deviation (measure (iii)) is also shown. For all six methods the averaged measures (ii) and (iv) are tabulated below, Tables 1 to 6.

4. Discussion

On the basis of the tabulated results, it can be observed that

TABLE 1. Turning Bands in Three Dimensions (Using 15 lines): Exponential Model

grid	range	average			standard deviation		
$L \times L$	$3a$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$
10×10	2.00	0.14	0.03	0.07	0.18	0.27	0.23
	5.00	0.04		0.04	0.25		0.25
	10.00			0.05			0.18
20×20	2.00	0.15	0.06	0.08	0.14	0.19	0.18
	5.00	0.03	0.01	0.02	0.15	0.25	0.20
	10.00	0.03		0.03	0.24		0.24
	15.00			0.02			0.19
	20.00			0.02			0.17
30×30	2.00	0.14	0.04	0.05	0.12	0.17	0.17
	5.00	0.04	0.05	0.05	0.12	0.19	0.16
	10.00	0.04	0.06	0.05	0.16	0.26	0.19
	15.00	0.03		0.03	0.17		0.17
	20.00			0.04			0.19
	30.00			0.02			0.16
60×60	2.00	0.13	0.04	0.05	0.07	0.10	0.10
	5.00	0.02	0.04	0.04	0.09	0.13	0.12
	10.00	0.01	0.03	0.02	0.11	0.18	0.16
	15.00	0.01	0.02	0.02	0.12	0.19	0.15
	20.00	0.02	0.04	0.03	0.13	0.24	0.16
	30.00	0.04		0.04	0.17		0.17
	40.00			0.04			0.19
	60.00			0.02			0.16

1) Turning Bands in three dimensions performs poorly for small effective ranges, characteristically having a higher variance than the other techniques and producing a poor fit for small effective ranges. From inspection of experimental semivariograms γ^* it was also seen that Turning Bands in three dimensions exhibits a tendency towards anisotropic behaviour, particularly when the range is small compared to the grid size.

2) Turning Bands in two dimensions produces preferable results to Turning Bands in three dimensions, in that it produces a better estimator of the mean and exhibits smaller variation between the individual semivariograms. Increasing the number of bands from 15 to 50 has the advantage of reducing the variability between the individual semivariograms as well as reducing the level of anisotropic behaviour

TABLE 2. LU Decomposition of the Covariance Matrix: Exponential Model

grid	range	average			standard deviation		
$L \times L$	$3a$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$
10×10	2	0.01	0.01	0.01	0.15	0.20	0.18
	5	0.02		0.02	0.21		0.21
	10			0.03			0.20
20×20	2	0.01	0.02	0.02	0.08	0.11	0.10
	5	0.02	0.01	0.01	0.09	0.19	0.14
	10	0.04		0.04	0.18		0.18
	15			0.01			0.17
	20			0.01			0.18
30×30	2	0.00	0.01	0.01	0.05	0.07	0.07
	5	0.01	0.01	0.01	0.07	0.13	0.11
	10	0.01	0.03	0.02	0.15	0.28	0.19
	15	0.02		0.02	0.19		0.19
	20			0.01			0.21
	30			0.02			0.21

observed within the experimental semivariograms γ^*. For very small effective ranges Turning Bands may be seen to break down (c.f. Tables 3 and 5 for an effective range of 2).

Mantoglou and Wilson [6] (page 1382) give an Abel integral equation which must be solved to find the one dimensional covariance structure. Importantly Brooker [1] (page 85) published an analytic solution to this equation, expressing the one dimensional covariance structure as a function of the two dimensional structure. High precision numerical approximations to this solution, in the case of the two dimensional exponential model, are readily obtained by standard techniques for numerical integration.

3) The LU, or Cholelsky, decomposition of the covariance matrix produced results which may be used as a benchmark for comparison of the other techniques. The LU decomposition produces data with the required structure, as opposed to the other techniques, which are approximations. Implementations of the LU decomposition are limited by the precision of the random number generator, and the accuracy of the machine arithmetic (Wilkinson [8]). Errorsare smaller than the errors due to approximation in the other techniques.

The LU, or Cholelsky, decomposition is not suitable for large grid sizes. To perform

TABLE 3. Turning Bands in Two Dimensions (Using 15 lines): Exponential Model

grid	range	average			standard deviation		
$L \times L$	$3a$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$
10×10	2	0.06	0.03	0.04	0.18	0.31	0.26
	5	0.03		0.03	0.19		0.19
	10			0.01			0.23
20×20	2	0.09	0.02	0.03	0.08	0.12	0.12
	5	0.04	0.02	0.03	0.12	0.24	0.18
	10	0.04		0.04	0.18		0.18
	15			0.02			0.18
	20			0.03			0.19
30×30	2	0.10	0.03	0.04	0.07	0.09	0.09
	5	0.02	0.02	0.02	0.07	0.13	0.11
	10	0.01	0.02	0.01	0.12	0.22	0.15
	15	0.01		0.01	0.17		0.17
	20			0.02			0.19
	30			0.03			0.16
60×60	2	0.10	0.03	0.03	0.05	0.06	0.06
	5	0.02	0.02	0.02	0.05	0.08	0.07
	10	0.01	0.02	0.01	0.07	0.15	0.12
	15	0.01	0.01	0.01	0.08	0.18	0.13
	20	0.03	0.05	0.04	0.13	0.25	0.17
	30	0.02		0.02	0.20		0.20
	40			0.01			0.17
	60			0.01			0.18

a simulation on L × L grid the LU decomposition uses memory of order L^4 and has a time complexity of order L^6. This is in contrast to the method of turning bands which has a time complexity and memory use of order L^2. Computers with greater than 2^{23} bytes of resident memory are readily available and it is a simple matter to produce 30 × 30 grids on such machines (Davis [3] recommended this as an upper bound in 1987.) It is possible to produce larger grids by the LU decomposition using larger machines, or secondary memory devices; the complexity and expense of doing so would tend to preclude this as a viable technique at the present stage.

    4) The Matrix Polynomial approximation performs well for small effective ranges.

TABLE 4. MPOLY, Square Root Decomposition of the Covariance Matrix (8 coefficients): Exponential Model

grid	range	average			standard deviation		
$L \times L$	$3a$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$
10×10	2	0.02	0.02	0.02	0.15	0.20	0.18
	5	0.03		0.03	0.21		0.21
	10			0.03			0.21
20×20	2	0.01	0.01	0.01	0.08	0.11	0.10
	5	0.03	0.02	0.02	0.09	0.19	0.14
	10	0.10		0.10	0.18		0.18
	15			0.03			0.18
	20			0.03			0.20
30×30	2	0.02	0.02	0.02	0.05	0.07	0.07
	5	0.03	0.02	0.02	0.07	0.14	0.11
	10	0.06	0.06	0.06	0.15	0.28	0.19
	15	0.02		0.02	0.18		0.18
	20			0.05			0.20
	30			0.16			0.21

A degradation of the reliability of the technique may be observed as the effective range grows, this being less pronounced for the 15 coefficient case than the 8.

Davis [3] proposed the use of a polynomial of degree 8, being the solution of a weighted minimax problem, but did not specify his weighting scheme. Here approximations to unweighted minimax polynomials with 8 and 15 coefficients were used. They were obtained by uniformly discretising the interval $[0, \lambda_{max}]$ and solving

$$\min_P \max_x |P(x) - \sqrt{x}|$$

upon the discretised domain, using the exchange algorithm. Here λ_{max} is the maximum eigenvalue of the covariance matrix, P is a polynomial of degree 7 or 14, and x is in the discretised interval. This weighting scheme gives a solution which is scale invariant, as opposed to that of Davis [3], which is not.

The Matrix Polynomial approximation has a memory complexity of order L^2 and a time complexity of order $L^2 \log(L)$ both of which compare favourably with the other techniques. It also complements the the techniques of TB3D and TB2D by producing good results for small effective ranges. For these reasons MPOLY and TB2D would be the recommended techniques, for the appropriate effective ranges.

TABLE 5. Turning Bands in Two Dimensions (Using 50 lines): Exponential Model

grid	range	average			standard deviation		
$L \times L$	$3a$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$
10×10	2	0.12	0.06	0.08	0.15	0.21	0.19
	5	0.03		0.03	0.22		0.22
	10			0.03			0.19
20×20	2	0.09	0.03	0.04	0.08	0.12	0.11
	5	0.03	0.03	0.03	0.10	0.18	0.14
	10	0.04		0.04	0.18		0.18
	15			0.01			0.20
	20			0.02			0.18
30×30	2	0.09	0.02	0.03	0.06	0.09	0.08
	5	0.03	0.03	0.03	0.06	0.12	0.10
	10	0.02	0.05	0.03	0.13	0.28	0.18
	15	0.02		0.02	0.16		0.16
	20			0.04			0.18
	30			0.01			0.16
60×60	2	0.10	0.02	0.02	0.03	0.05	0.05
	5	0.02	0.01	0.01	0.04	0.07	0.06
	10	0.01	0.01	0.01	0.06	0.13	0.11
	15	0.01	0.02	0.01	0.08	0.18	0.13
	20	0.02	0.05	0.03	0.12	0.22	0.15
	30	0.02		0.02	0.15		0.15
	40			0.01			0.15
	60			0.03			0.17

5) The amount of variation from the averaged semivariogram may be seen to grow as the lag distances increase.

6) As the grid size increases, semivariograms of a specified range are better defined.

7) With the exception discussed above for very small effective ranges with the turning bands algorithms, it appears that for a fixed grid size, $L \times L$, data sets for semivariograms with smaller ranges are more accurately simulated.

8) It is worth emphasising that the statistics tabulated above were calculated from 50 grids, and are subject to statistical fluctuation. Checks on ensembles of 200 grids for several methods indicate that the results using 50 grids were reliable.

TABLE 6. MPOLY, Square Root Decomposition of the Covariance Matrix (15 coefficients): Exponential Model

grid	range	average			standard deviation		
$L \times L$	$3a$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$	$[1, 3a]$	$(3a, L/2]$	$[1, L/2]$
10×10	2	0.01	0.01	0.01	0.15	0.20	0.18
	5	0.03		0.03	0.21		0.21
	10			0.03			0.21
20×20	2	0.01	0.01	0.01	0.08	0.11	0.10
	5	0.01	0.02	0.02	0.09	0.19	0.14
	10	0.06		0.06	0.18		0.18
	15			0.03			0.18
	20			0.02			0.20
30×30	2	0.00	0.01	0.01	0.05	0.07	0.07
	5	0.01	0.01	0.01	0.07	0.13	0.11
	10	0.03	0.03	0.03	0.14	0.27	0.19
	15	0.02		0.02	0.19		0.19
	20			0.01			0.20
	30			0.04			0.21

5. Acknowledgement

References

[1] P. I. Brooker, "Two-dimensional simulation by turning bands", *Mathematical Geology* **17** (1985) 81–90.

[2] P. I. Brooker, G. C. Cock and M. A. Stewart, "Comparison of methods for simulation of two dimensional data honouring specified spherical semivariograms", *Mathematics and Computers in Simulation* **33** (1992) 489–494.

[3] M. W. Davis, "Generating large stochastic simulations - the matrix polynomial approximation method", *Mathematical Geology* **19** (1987) 99–108.

[4] M. W. Davis, "Production of conditional simulations via the lu triangular decomposition of the covariance matrix", *Mathematical Geology* **19** (1987) 91–98.

[5] A. G. Journel and C. J. Huijbregts, *Mining geostatistics* (Academic Press, London, 1978).

[6] A. Mantoglou and J. L. Wilson, "The turning bands method for simulation of random fields using line generation by spectral method", *Water Resources Research* **18** (1982) 1379–1394.

[7] A. F. B. Tompson, R. Ababou and L. W. Gelhar, "Implementation of the three-dimensional turning bands random field generator", *Water Resources Research* **25** (1989) 2227–2243.

[8] J. H. Wilkinson, "A priori error analyses of algebraic processes", *Proceedings of the International Congress of Mathematicians* (1968) 629–631.