Durham University

Durham E-Theses

Error processes in the integration of digital cartographic data in geographic information systems.

Krysia Rybaczuk

How to cite:

Rybaczuk, Krysia (1992) Error processes in the integration of digital cartographic data in geographic information systems. Doctoral thesis, Durham University.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a https://etheses.durham.ac.uk/id/eprint/6210/ is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
https://etheses.durham.ac.uk

Error processes in the integration of digital cartographic data in geographic information systems.

Krysia Rybaczuk

Phd Thesis

1991

Volume 2

The copyright of this thesis rests with the author.

DURHAM UNIVERSITY

- 9 JUL 1992

Volume 2 - Figures

Table of Figures held in Volume Two

Chapter Two

2. 1. The POLVRT data structure
2. 2. The topological data structure as used by GIMMS

Chapter Three

3. 1. A model of error transmission in a GIS
3. 2. Common methods of cartographic generalisation

Chapter Four

4. 1. Components of digitising
4. 2. Stream digitising
4. 3. Costs and benefits of three data capture methodologies
4. 4. Error components of the table digitising process
4. 5. Problems of line following when digitising
4. 6. Variations in the sum of error for the entire grid
4. 7. Variations in the total deviation for individual points
4. 8. Mean and mode values for operators in the grid experiment
4. 9. A frequency distribution of the total error encountered by operator MJB
4. 10. A frequency distribution of the total error encountered by operator ALS
4. 11. A frequency distribution of the total error encountered by operator CEH
4. 12. A frequency distribution of the total error encountered by operator NJP
4. 13. A frequency distribution of the total error encountered by operator KYR
4. 14. A frequency distribution of the total error encountered by operator GLG
4. 15. Variations in the X and Y coordinates achieved by MJB whilst digitising
4. 16. Variations in the X and Y coordinates achieved by ALS whilst digitising
4. 17. Variations in the X and Y coordinates achieved by CEH whilst digitising
4. 18. Variations in the X and Y coordinates achieved by NJP whilst digitising
4. 19. Variations in the X and Y coordinates achieved by KYR whilst digitising
4. 20. Variations in the X and Y coordinates achieved by GLG whilst digitising
4. 21. Variations in the X coordinate whilst digitising = 900 points digitised by MJB
4. 22. Variations in the X coordinate whilst digitising = 900 points digitised by ALS
4. 23. Variations in the X coordinate whilst digitising = 900 points digitised by CEH
4. 24. Variations in the X coordinate whilst digitising = 900 points digitised by NJP
4. 25. Variations in the X coordinate whilst digitising = 900 points digitised by KYR
4. 26. Variations in the X coordinate whilst digitising = 900 points digitised by GLG
4. 27. Variations in the Y coordinate whilst digitising = 900 points digitised by MJB
4. 28. Variations in the Y coordinate whilst digitising = 900 points digitised by ALS
4. 29. Variations in the Y coordinate whilst digitising = 900 points digitised by CEH
4. 30. Variations in the Y coordinate whilst digitising = 900 points digitised by NJP
4. 31. Variations in the Y coordinate whilst digitising = 900 points digitised by KYR
4. 32. Variations in the Y coordinate whilst digitising = 900 points digitised by GLG
4. 33. Variations in the X coordinate whilst digitising = 100 points digitised by GLG
4. 34. Variations in the Y coordinate whilst digitising = 100 points digitised by GLG
4. 35. Variations in the X coordinate whilst digitising = 100 points digitised under timed conditions by GLG

4. 36. Variations in the Y coordinate whilst digitising = 100 points digitised under timed conditions by GLG
4. 37. Digitising error with respect to the spatial extent of the digitising table for operator NJP
4. 38. Digitising error with respect to the spatial extent of the digitising table for operator CEH
4. 39. Digitising error with respect to the spatial extent of the digitising table for operator KYR
4. 40. Digitising error with respect to the spatial extent of the digitising table for operator MJB
4. 41. Digitising error with respect to the spatial extent of the digitising table for operator GLG
4. 42. Digitising error with respect to the spatial extent of the digitising table for operator ALS
4. 43. A frequency histogram of the length of lines when digitised from different scales using different operators
4. 44. Variations in the length of line with respect to points
4. 45. Confidence limits for the similarity of digitised lines
4. 46. Using too many points to efficiently characterise a line
4. 47. A comparison of the same feature with differing lengths digitised by PNA from the 1:25,000 scale map

Chapter Five

5. 1. Network structured data illustrating the need for contiguity so that flows can be modelled
5. 2. Data model unable to create full polygon structure due to the presence of a dangling arc
5. 3. Snap tolerance used in digitising
5. 4. Entity by entity digitising and possible problems
5. 5. Priority numbering of polygons as an alternative to have digitise complex lines twice
5. 6. Undershoots and overshoots
5. 7a. Vertex move and associated problems
5. 7b. Arc move and associated problems
5. 8. Problems of vertex movement using straight line coverages
5. 9. Problems of node collapse
5. 10. The unpredictability of node matches as a consequence of random node numbering
5. 11. The removal of intentional 'dangling' lines in an attempt to remove node errors
5. 12. The weeding procedure
5. 13. Original coverage: digitised with both fuzzy and dangle tol = 5 cov. units
5. 14. Fuzzy tol = 50 coverage units: dangle tol = 60 coverage units
5. 15. Mnode tol = 30 cov units: fuzzy tol = 5 cov units: dangle tol = 30 cov units
5. 16. Fuzzy tol = 50 cov units: dangle tol = 30 cov units
5. 17. Mnode = 30 coverage units
5. 18. Mnode = 50 coverage units
5. 19. Mnode tol = 50 cov units: fuzzy tol = 5 cov units: dangle tol = 60 cov units
5. 20. Mnode tol = 30 cov units: fuzzy tol = 5 cov units: dangle tol = 60 cov units
5. 21. Fuzzy tol = 8 coverage units: dangle tol = 60 coverage units
5. 22. Mnode tol = 3 cov units: fuzzy tol = 5 cov units: dangle tol = 2 cov units

5. 23. Mnode tol = 5 cov units: fuzzy tol = 5 cov units: dangle tol = 60 cov units
5. 24. Mnode tol = 50 cov units: fuzzy tol = 5 cov units: dangle tol = 30 cov units
5. 25. Fuzzy tol = 30 coverage units: dangle tol = 8 coverage units
5. 26. Mnode = 20 coverage units
5. 27. Mnode = 8 coverage units
5. 28. Mnode = 5 coverage units
5. 29. Mnode = 3 coverage units
5. 30. Mnode tol = 3 cov units: fuzzy tol = 5 cov units: dangle tol = 60 cov units
5. 31. Mnode tol = 3 cov units: fuzzy tol = 5 cov units: dangle tol = 30 cov units
5. 32. Mnode tol = 3 cov units: fuzzy tol = 5 cov units: dangle tol = 8 cov units
5. 33. Mnode tol = 8 cov units: fuzzy tol = 5 cov units: dangle tol = 2 cov units
5. 34. Mnode tol = 8 cov units: fuzzy tol = 5 cov units: dangle tol = 8 cov units
5. 35. Mnode tol = 8 cov units: fuzzy tol = 5 cov units: dangle tol = 30 cov units
5. 36. Mnode tol = 8 cov units: fuzzy tol = 5 cov units: dangle tol = 60 cov units
5. 37. Fuzzy tolerance = 3 coverage units
5. 38. Fuzzy tolerance = 3 coverage units: dangle tol = 2 coverage units
5. 39. Fuzzy tol = 3 coverage units: dangle tol = 30 coverage units
5. 40. Fuzzy tol = 5 coverage units: dangle tol = 2 coverage units
5. 41. Fuzzy tol = 5 coverage units: dangle tol = 30 coverage units
5. 42. Fuzzy tol = 8 coverage units: dangle tol = 2 coverage units
5. 43. Fuzzy tol = 8 coverage units: dangle tol = 30 coverage units
5. 44. Fuzzy tol = 30 coverage units: dangle tol = 2 coverage units
5. 45. Fuzzy tol = 30 coverage units: dangle tol = 30 coverage units
5. 46. Fuzzy tol = 30 coverage units: dangle tol = 60 coverage units
5. 47. Subset of fenland coverage used for subsequent experiments
5. 48. Number of errors removed by various node matching tolerances
5. 49. Errors versus lines for the match node tolerance
5. 50. Problems of dangle removal following the use of the intersection program
5. 51. The distribution of the total number of errors in the coverages following processing
5. 52. The distribution of the total number of lines in coverages following processing
5. 53. The percentage of lines remaining with respect to the lines removed
5. 54. A plot of the total number of errors retained as compared with the total number of lines retained
5. 55. Histograms of the line distributions with respect to the type and size of tolerance used
5. 56. Lines less than 0.009in retained in the coverage
5. 57. Lines greater than 0.009in and less than 0.01in retained in the coverage
5. 58. Lines greater than 0.01in and less than 0.02in retained in the coverage
5. 59. Lines greater than 0.02in and less than 0.03in retained in the coverage
5. 60. Lines greater than 0.03in and less than 0.05in retained in the coverage
5. 61. Lines greater than 0.05in and less than 0.09in retained in the coverage
5. 62. Lines greater than 0.09in and less than 0.2in retained in the coverage
5. 63. Lines greater than 0.2in retained in the coverage
5. 64. Distribution of coverage line lengths following a match tol of 0.009" and a dangle tol of 0.03"
5. 65. Distribution of coverage line lengths following a match of 0.009" and a dangle tol of 0.2"
5. 66. Distribution of coverage line lengths following a match tol of 0.09" and a dangle tol of 0.03"

5. 67. Distribution of coverage line lengths following a match tol of 0.026" and a dangle tol of 0.02"
5. 68. Distribution of coverage line lengths following a match tol of 0.026" and a dangle tol of 0.2"
5. 69. Distribution of coverage line lengths following a match tol of 0.09" and a dangle tol of 0.2"
5. 70. Distribution of coverage line lengths following a match tol of 0.009", a dangle tol of 0.2" and a weed tol of 0.1"
5. 71. Distribution of coverage line lengths following a match tol of 0.09", a dangle tol of 0.2" and a weed tol of 0.1"
5. 72. A frequency distribution of the lengths T02602 coverage following a square root transformation on the data
5. 73. A frequency distribution of the lengths in the T0902W1 coverage following a square root transformation on the data
5. 74. A comparison of T-test results.
5. 75. Reduction of polygon size due to poor digitising
5. 76. The destructive power of faulty intersecting
5. 77. The use of an additional identifier to track lines in ARC/INFO
5. 78. Evaluating the changes in line length following tolerancing
5. 79. Problems of line lengths and geometric movements
5. 80. Linear collapse and its effects on line identifiers
5. 81. Distibtuion of node movement for coverages T026c02, T026c2, T09c03 and T09c2
5. 82. Distibtuion of node movement for coverages T09c2w1, T009c03, T009c2 and T009c2w1
5. 83. Subset of original coverage showing line numbers and node errors
5. 84. Subset of T009C03 coverage showing line numbers and node errors
5. 85. Subset of T009C2 coverage showing line numbers and node errors
5. 86. Subset of T009C2W1 coverage showing line numbers and node errors
5. 87. Perspectives on geometric change as tolerance values are increased

Chapter Six

6. 1. The nature of set theory adoption in ARC/INFO
6. 2. Imperfect matching in overlay due to the variability of source data
6. 3. A tracing of 1:25,000 Ordnance Survey map used as input showing feature generalisation
6. 4. The transformation of dual lines onto single entities
6. 5. A section of the 1:50,000 Ordnance Survey map used as the basis for coverage input
6. 6. Problems of digitising lines represented as areas in the source document
6. 7. Problems of the bouncing line during digitising
6. 8. Problems of topology as a consequence of digitising errors
6. 9. Options for manual editing of geometric discrepancies
6. 10. Diagram showing the division of the full map into eight sub-sections
6. 11. Map of the Fenlands digitised from the 1:10,560 scale map (b)
6. 12. Map of the Fenlands digitised from the 1:10,560 scale map (a)
6. 13. Map of the Fenlands digitised from the 1:25,000 scale map
6. 14. Map of the Fenlands digitised from the 1:50,000 scale map (a)
6. 15. Map of the Fenlands digitised from the 1:50,000 scale map (b)
6. 16. Overlay of the two 1:50,000 scale coverages

6. 17. Section one of overlaid coverages
6. 18. Section two of overlaid coverages
6. 19. Section three of overlaid coverages
6. 20. Section four of overlaid coverages
6. 21. Section five of overlaid coverages
6. 22. Section six of overlaid coverages
6. 23. Section seven of overlaid coverages
6. 24. Section eight of overlaid coverages
6. 25. A frequency histogram of polygon size for the 1:50,000b coverage
6. 26. A frequency histogram of polygon size for the overlay of the two 1:50,000 coverages
6. 27. A frequency distribution of polygon size in the 1:10,560b coverage
6. 28. A frequency distribution of polygon size in the overlaid coverage consisting of two versions of the 1:10,560 Ordnance Survey maps
6. 29. A frequency histogram showing the distribution of polygon size for the 1:25,000 coverage
6. 30. A frequency histogram showing the distribution of polygon size following an overlay of the 1:50,000b, 1:25,000 and 1:10,560 coverages
6. 31. A frequency histogram of polygon size following an overlay of the 1:10,560b coverage and the 1:25,000 coverage
6. 32. Frequency histogram of the two overlays involving the 1:50,000b coverage and the coverages of other scales
6. 33. The nature of sliver polygons
6. 34. Area as an unreliable parameter for sliver polygon removal
6. 35. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of PERIMETER = 1,000 units
6. 36. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of PERIMETER = 2,000 units
6. 37. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of PERIMETER = 2,500 units
6. 38. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA = 5,000 units
6. 39. An overlay of the 1:50,000a coverage and the 1,50,000b coverage following an ELIMINATE of AREA = 8,000 units
6. 40. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA = 150,000 units
6. 41. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA = 50,000 units
6. 42. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 10 units
6. 43. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 12 units
6. 44. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 15 units
6. 45. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 20 units
6. 46. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 25 units
6. 47. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an

ELIMINATE of AREA/PERIMETER = 30 units
6. 48. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 50 units
6. 49. An overlay of the 1:50,000b coverage, the 1:25,000 coverage and the 1:10,560b coverage following an ELIMINATE of AREA/PERIMETER = 10 units
6. 50. An overlay of the 1:50,000b coverage, the 1:25,000 coverage and the 1:10,560b coverage following an ELIMINATE of AREA/PERIMETER = 20 units
6. 51. An overlay of the 1:50,000b coverage, the 1:25,000 coverage and the 1:10,560b coverage following an ELIMINATE of AREA/PERIMETER = 30 units

Chapter Seven

7. 1. Using polygon identifiers as a basis for overlay error detection
7. 2. A test coverage consisting of small zones and long thin zones
7. 3. The test coverages showing assigned land use values
7. 4. Overlay of the original coverage and the transformed coverages each consisting of the same features
7. 5. Coverage of change
7. 6. Overlay of the original coverage and the coverage of change
7. 7. ELIMINATION of errors by forcing coverages to match on the basis of equal identifiers
7. 8. ELIMINATION on the basis of AREA/PERIMETER being less than 0.1 units
7. 9. Determining allowable movement on the basis of apriori knowledge of the data
7. 10. Using CLASS as the basis for ELIMINATION
7. 11. Using either CLASS or CLASS2 as the basis for ELIMINATION
7. 12. Dissolving on the CLASS2 attribute following the use of CLASS or CLASS2 as the basis for ELIMINATION
7. 13. ELIMINATION on the basis of AREA/PERIMETER being less than 10 units
7. 14. ELIMINATION on the basis of AREA/PERIMETER being less than 0.5 units
7. 15. ELIMINATION on the basis of PERIMETER/AREA being greater than 10 units
7. 16. Coverage dissolved on the basis of database rules without the use of the ELIMINATE option
7. 17. A methodology for selective neighbourhood matching for sliver polygons
7. 18. Problems of sequential sliver processing
7. 19. Land use categories assigned to the first coverage
7. 20. Land use categories assigned to the overlay coverage
7. 21. The overlaid coverage showing their histories
7. 22. The overlaid coverages showing changed histories as defined by the matrix of probable histories
7. 23. The overlaid coverages following a DISSOLVE operation on the new polygon history values.

Table of Figures held in Volume Two

Chapter Two

2. 1. The POLVRT data structure
2. 2. The topological data structure as used by GIMMS

Chapter Three

3. 1. A model of error transmission in a GIS
3. 2. Common methods of cartographic generalisation

Chapter Four

4. 1. Components of digitising
4. 2. Stream digitising
4. 3. Costs and benefits of three data capture methodologies
4. 4. Error components of the table digitising process
4. 5. Problems of line following when digitising
4. 6. Variations in the sum of error for the entire grid
4. 7. Variations in the total deviation for individual points
4. 8. Mean and mode values for operators in the grid experiment
4. 9. A frequency distribution of the total error encountered by operator MJB
4. 10. A frequency distribution of the total error encountered by operator ALS
4. 11. A frequency distribution of the total error encountered by operator CEH
4. 12. A frequency distribution of the total error encountered by operator NJP
4. 13. A frequency distribution of the total error encountered by operator KYR
4. 14. A frequency distribution of the total error encountered by operator GLG
4. 15. Variations in the X and Y coordinates achieved by MJB whilst digitising
4. 16. Variations in the X and Y coordinates achieved by ALS whilst digitising
4. 17. Variations in the X and Y coordinates achieved by CEH whilst digitising
4. 18. Variations in the X and Y coordinates achieved by NJP whilst digitising
4. 19. Variations in the X and Y coordinates achieved by KYR whilst digitising
4. 20. Variations in the X and Y coordinates achieved by GLG whilst digitising
4. 21. Variations in the X coordinate whilst digitising = 900 points digitised by MJB
4. 22. Variations in the X coordinate whilst digitising = 900 points digitised by ALS
4. 23. Variations in the X coordinate whilst digitising = 900 points digitised by CEH
4. 24. Variations in the X coordinate whilst digitising = 900 points digitised by NJP
4. 25. Variations in the X coordinate whilst digitising = 900 points digitised by KYR
4. 26. Variations in the X coordinate whilst digitising = 900 points digitised by GLG
4. 27. Variations in the Y coordinate whilst digitising = 900 points digitised by MJB
4. 28. Variations in the Y coordinate whilst digitising = 900 points digitised by ALS
4. 29. Variations in the Y coordinate whilst digitising = 900 points digitised by CEH
4. 30. Variations in the Y coordinate whilst digitising = 900 points digitised by NJP
4. 31. Variations in the Y coordinate whilst digitising = 900 points digitised by KYR
4. 32. Variations in the Y coordinate whilst digitising = 900 points digitised by GLG
4. 33. Variations in the X coordinate whilst digitising = 100 points digitised by GLG
4. 34. Variations in the Y coordinate whilst digitising = 100 points digitised by GLG
4. 35. Variations in the X coordinate whilst digitising = 100 points digitised under timed conditions by GLG

4. 36. Variations in the Y coordinate whilst digitising = 100 points digitised under timed conditions by GLG
4. 37. Digitising error with respect to the spatial extent of the digitising table for operator NJP
4. 38. Digitising error with respect to the spatial extent of the digitising table for operator CEH
4. 39. Digitising error with respect to the spatial extent of the digitising table for operator KYR
4. 40. Digitising error with respect to the spatial extent of the digitising table for operator MJB
4. 41. Digitising error with respect to the spatial extent of the digitising table for operator GLG
4. 42. Digitising error with respect to the spatial extent of the digitising table for operator ALS
4. 43. A frequency histogram of the length of lines when digitised from different scales using different operators
4. 44. Variations in the length of line with respect to points
4. 45. Confidence limits for the similarity of digitised lines
4. 46. Using too many points to efficiently characterise a line
4. 47. A comparison of the same feature with differing lengths digitised by PNA from the 1:25,000 scale map

Chapter Five

5. 1. Network structured data illustrating the need for contiguity so that flows can be modelled
5. 2. Data model unable to create full polygon structure due to the presence of a dangling arc
5. 3. Snap tolerance used in digitising
5. 4. Entity by entity digitising and possible problems
5. 5. Priority numbering of polygons as an alternative to have digitise complex lines twice
5. 6. Undershoots and overshoots
5. 7a. Vertex move and associated problems
5. 7b. Arc move and associated problems
5. 8. Problems of vertex movement using straight line coverages
5. 9. Problems of node collapse
5. 10. The unpredictability of node matches as a consequence of random node numbering
5. 11. The removal of intentional 'dangling' lines in an attempt to remove node errors
5. 12. The weeding procedure
5. 13. Original coverage: digitised with both fuzzy and dangle tol = 5 cov. units
5. 14. Fuzzy tol = 50 coverage units: dangle tol = 60 coverage units
5. 15. Mnode tol = 30 cov units: fuzzy tol = 5 cov units: dangle tol = 30 cov units
5. 16. Fuzzy tol = 50 cov units: dangle tol = 30 cov units
5. 17. Mnode = 30 coverage units
5. 18. Mnode = 50 coverage units
5. 19. Mnode tol = 50 cov units: fuzzy tol = 5 cov units: dangle tol = 60 cov units
5. 20. Mnode tol = 30 cov units: fuzzy tol = 5 cov units: dangle tol = 60 cov units
5. 21. Fuzzy tol = 8 coverage units: dangle tol = 60 coverage units
5. 22. Mnode tol = 3 cov units: fuzzy tol = 5 cov units: dangle tol = 2 cov units

5. 23. Mnode tol = 5 cov units: fuzzy tol = 5 cov units: dangle tol = 60 cov units
5. 24. Mnode tol = 50 cov units: fuzzy tol = 5 cov units: dangle tol = 30 cov units
5. 25. Fuzzy tol = 30 coverage units: dangle tol = 8 coverage units
5. 26. Mnode = 20 coverage units
5. 27. Mnode = 8 coverage units
5. 28. Mnode = 5 coverage units
5. 29. Mnode = 3 coverage units
5. 30. Mnode tol = 3 cov units: fuzzy tol = 5 cov units: dangle tol = 60 cov units
5. 31. Mnode tol = 3 cov units: fuzzy tol = 5 cov units: dangle tol = 30 cov units
5. 32. Mnode tol = 3 cov units: fuzzy tol = 5 cov units: dangle tol = 8 cov units
5. 33. Mnode tol = 8 cov units: fuzzy tol = 5 cov units: dangle tol = 2 cov units
5. 34. Mnode tol = 8 cov units: fuzzy tol = 5 cov units: dangle tol = 8 cov units
5. 35. Mnode tol = 8 cov units: fuzzy tol = 5 cov units: dangle tol = 30 cov units
5. 36. Mnode tol = 8 cov units: fuzzy tol = 5 cov units: dangle tol = 60 cov units
5. 37. Fuzzy tolerance = 3 coverage units
5. 38. Fuzzy tolerance = 3 coverage units: dangle tol = 2 coverage units
5. 39. Fuzzy tol = 3 coverage units: dangle tol = 30 coverage units
5. 40. Fuzzy tol = 5 coverage units: dangle tol = 2 coverage units
5. 41. Fuzzy tol = 5 coverage units: dangle tol = 30 coverage units
5. 42. Fuzzy tol = 8 coverage units: dangle tol = 2 coverage units
5. 43. Fuzzy tol = 8 coverage units: dangle tol = 30 coverage units
5. 44. Fuzzy tol = 30 coverage units: dangle tol = 2 coverage units
5. 45. Fuzzy tol = 30 coverage units: dangle tol = 30 coverage units
5. 46. Fuzzy tol = 30 coverage units: dangle tol = 60 coverage units
5. 47. Subset of fenland coverage used for subsequent experiments
5. 48. Number of errors removed by various node matching tolerances
5. 49. Errors versus lines for the match node tolerance
5. 50. Problems of dangle removal following the use of the intersection program
5. 51. The distribution of the total number of errors in the coverages following processing
5. 52. The distribution of the total number of lines in coverages following processing
5. 53. The percentage of lines remaining with respect to the lines removed
5. 54. A plot of the total number of errors retained as compared with the total number of lines retained
5. 55. Histograms of the line distributions with respect to the type and size of tolerance used
5. 56. Lines less than 0.009in retained in the coverage
5. 57. Lines greater than 0.009in and less than 0.01in retained in the coverage
5. 58. Lines greater than 0.01in and less than 0.02in retained in the coverage
5. 59. Lines greater than 0.02in and less than 0.03in retained in the coverage
5. 60. Lines greater than 0.03in and less than 0.05in retained in the coverage
5. 61. Lines greater than 0.05in and less than 0.09in retained in the coverage
5. 62. Lines greater than 0.09in and less than 0.2in retained in the coverage
5. 63. Lines greater than 0.2in retained in the coverage
5. 64. Distribution of coverage line lengths following a match tol of 0.009" and a dangle tol of 0.03"
5. 65. Distribution of coverage line lengths following a match of 0.009" and a dangle tol of 0.2"
5. 66. Distribution of coverage line lengths following a match tol of 0.09" and a dangle tol of 0.03"

5. 67. Distribution of coverage line lengths following a match tol of 0.026" and a dangle tol of 0.02"
5. 68. Distribution of coverage line lengths following a match tol of 0.026" and a dangle tol of 0.2"
5. 69. Distribution of coverage line lengths following a match tol of 0.09" and a dangle tol of 0.2"
5. 70. Distribution of coverage line lengths following a match tol of 0.009", a dangle tol of 0.2" and a weed tol of 0.1"
5. 71. Distribution of coverage line lengths following a match tol of 0.09", a dangle tol of 0.2" and a weed tol of 0.1"
5. 72. A frequency distribution of the lengths T02602 coverage following a square root transformation on the data
5. 73. A frequency distribution of the lengths in the T0902W1 coverage following a square root transformation on the data
5. 74. A comparison of T-test results.
5. 75. Reduction of polygon size due to poor digitising
5. 76. The destructive power of faulty intersecting
5. 77. The use of an additional identifier to track lines in ARC/INFO
5. 78. Evaluating the changes in line length following tolerancing
5. 79. Problems of line lengths and geometric movements
5. 80. Linear collapse and its effects on line identifiers
5. 81. Distibtuion of node movement for coverages T026c02, T026c2, T09c03 and T09c2
5. 82. Distibtuion of node movement for coverages T09c2w1, T009c03, T009c2 and T009c2w1
5. 83. Subset of original coverage showing line numbers and node errors
5. 84. Subset of T009C03 coverage showing line numbers and node errors
5. 85. Subset of T009C2 coverage showing line numbers and node errors
5. 86. Subset of T009C2W1 coverage showing line numbers and node errors
5. 87. Perspectives on geometric change as tolerance values are increased

Chapter Six

6. 1. The nature of set theory adoption in ARC/INFO
6. 2. Imperfect matching in overlay due to the variability of source data
6. 3. A tracing of 1:25,000 Ordnance Survey map used as input showing feature generalisation
6. 4. The transformation of dual lines onto single entities
6. 5. A section of the 1:50,000 Ordnance Survey map used as the basis for coverage input
6. 6. Problems of digitising lines represented as areas in the source document
6. 7. Problems of the bouncing line during digitising
6. 8. Problems of topology as a consequence of digitising errors
6. 9. Options for manual editing of geometric discrepancies
6. 10. Diagram showing the division of the full map into eight sub-sections
6. 11. Map of the Fenlands digitised from the 1:10,560 scale map (b)
6. 12. Map of the Fenlands digitised from the 1:10,560 scale map (a)
6. 13. Map of the Fenlands digitised from the 1:25,000 scale map
6. 14. Map of the Fenlands digitised from the 1:50,000 scale map (a)
6. 15. Map of the Fenlands digitised from the 1:50,000 scale map (b)
6. 16. Overlay of the two 1:50,000 scale coverages

6. 17. Section one of overlaid coverages
6. 18. Section two of overlaid coverages
6. 19. Section three of overlaid coverages
6. 20. Section four of overlaid coverages
6. 21. Section five of overlaid coverages
6. 22. Section six of overlaid coverages
6. 23. Section seven of overlaid coverages
6. 24. Section eight of overlaid coverages
6. 25. A frequency histogram of polygon size for the 1:50,000b coverage
6. 26. A frequency histogram of polygon size for the overlay of the two 1:50,000 coverages
6. 27. A frequency distribution of polygon size in the 1:10,560b coverage
6. 28. A frequency distribution of polygon size in the overlaid coverage consisting of two versions of the 1:10,560 Ordnance Survey maps
6. 29. A frequency histogram showing the distribution of polygon size for the 1:25,000 coverage
6. 30. A frequency histogram showing the distribution of polygon size following an overlay of the 1:50,000b, 1:25,000 and 1:10,560 coverages
6. 31. A frequency histogram of polygon size following an overlay of the 1:10,560b coverage and the 1:25,000 coverage
6. 32. Frequency histogram of the two overlays involving the 1:50,000b coverage and the coverages of other scales
6. 33. The nature of sliver polygons
6. 34. Area as an unreliable parameter for sliver polygon removal
6. 35. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of PERIMETER = 1,000 units
6. 36. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of PERIMETER = 2,000 units
6. 37. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of PERIMETER = 2,500 units
6. 38. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA = 5,000 units
6. 39. An overlay of the 1:50,000a coverage and the 1,50,000b coverage following an ELIMINATE of AREA = 8,000 units
6. 40. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA = 150,000 units
6. 41. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA = 50,000 units
6. 42. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 10 units
6. 43. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 12 units
6. 44. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 15 units
6. 45. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 20 units
6. 46. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 25 units
6. 47. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an

ELIMINATE of AREA/PERIMETER = 30 units
6. 48. An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER = 50 units
6. 49. An overlay of the 1:50,000b coverage, the 1:25,000 coverage and the 1:10,560b coverage following an ELIMINATE of AREA/PERIMETER = 10 units
6. 50. An overlay of the 1:50,000b coverage, the 1:25,000 coverage and the 1:10,560b coverage following an ELIMINATE of AREA/PERIMETER = 20 units
6. 51. An overlay of the 1:50,000b coverage, the 1:25,000 coverage and the 1:10,560b coverage following an ELIMINATE of AREA/PERIMETER = 30 units

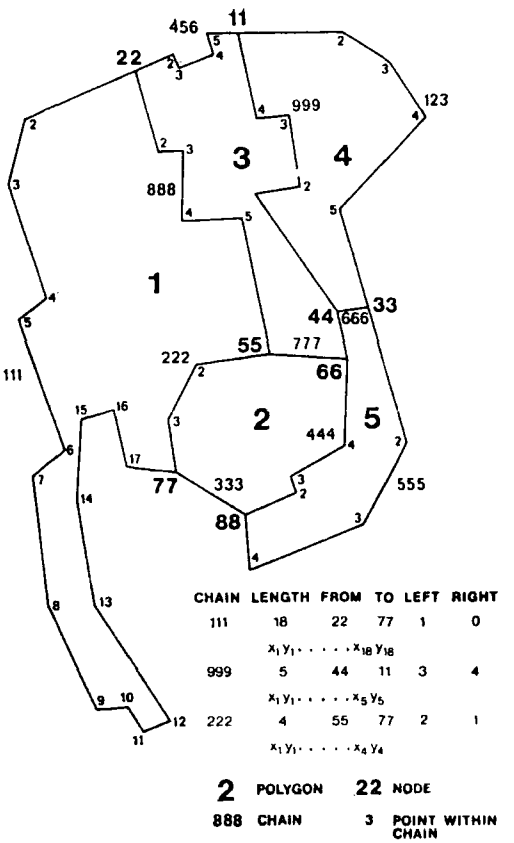
Chapter Seven

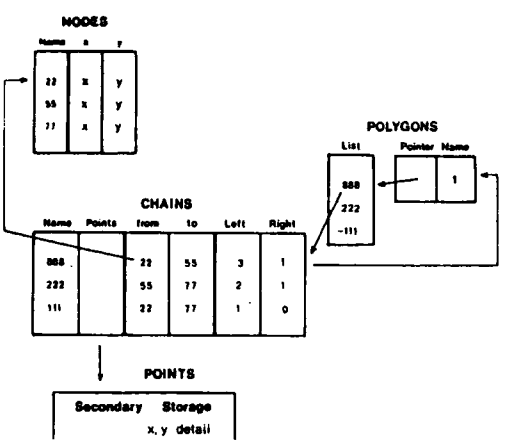
7. 1. Using polygon identifiers as a basis for overlay error detection
7. 2. A test coverage consisting of small zones and long thin zones
7. 3. The test coverages showing assigned land use values
7. 4. Overlay of the original coverage and the transformed coverages each consisting of the same features
7. 5. Coverage of change
7. 6. Overlay of the original coverage and the coverage of change
7. 7. ELIMINATION of errors by forcing coverages to match on the basis of equal identifiers
7. 8. ELIMINATION on the basis of AREA/PERIMETER being less than 0.1 units
7. 9. Determining allowable movement on the basis of apriori knowledge of the data
7. 10. Using CLASS as the basis for ELIMINATION
7. 11. Using either CLASS or CLASS2 as the basis for ELIMINATION
7. 12. Dissolving on the CLASS2 attribute following the use of CLASS or CLASS2 as the basis for ELIMINATION
7. 13. ELIMINATION on the basis of AREA/PERIMETER being less than 10 units
7. 14. ELIMINATION on the basis of AREA/PERIMETER being less than 0.5 units
7. 15. ELIMINATION on the basis of PERIMETER/AREA being greater than 10 units
7. 16. Coverage dissolved on the basis of database rules without the use of the ELIMINATE option
7. 17. A methodology for selective neighbourhood matching for sliver polygons
7. 18. Problems of sequential sliver processing
7. 19. Land use categories assigned to the first coverage
7. 20. Land use categories assigned to the overlay coverage
7. 21. The overlaid coverage showing their histories
7. 22. The overlaid coverages showing changed histories as defined by the matrix of probable histories
7. 23. The overlaid coverages following a DISSOLVE operation on the new polygon history values.

Figures for Chapter 2

UNIV DURHAM

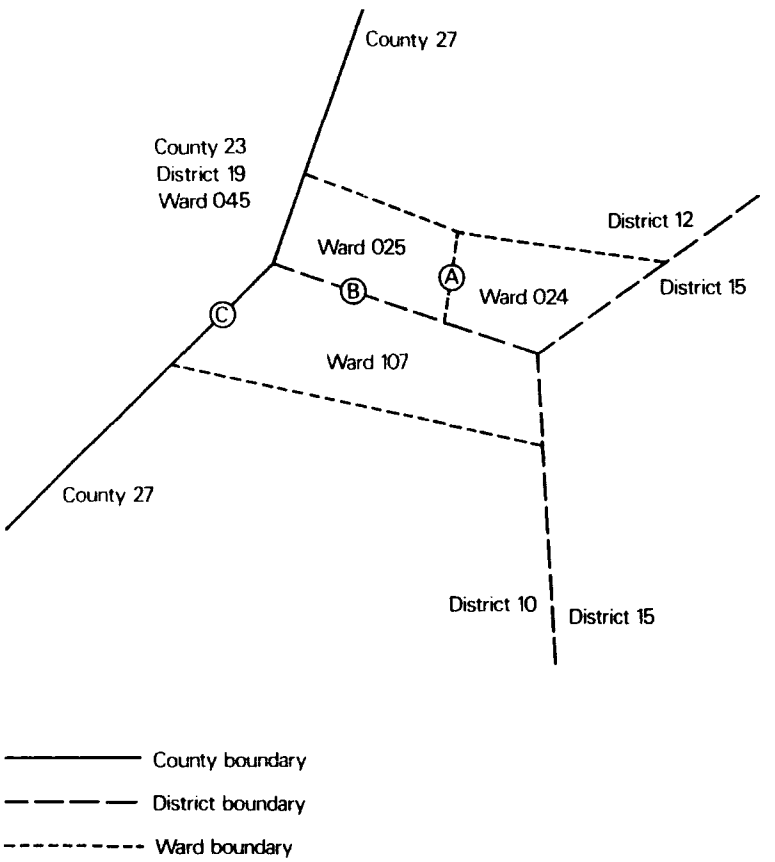
2.1 The POLYVRT data structure





Source: Chrisman and Peucker (1975)

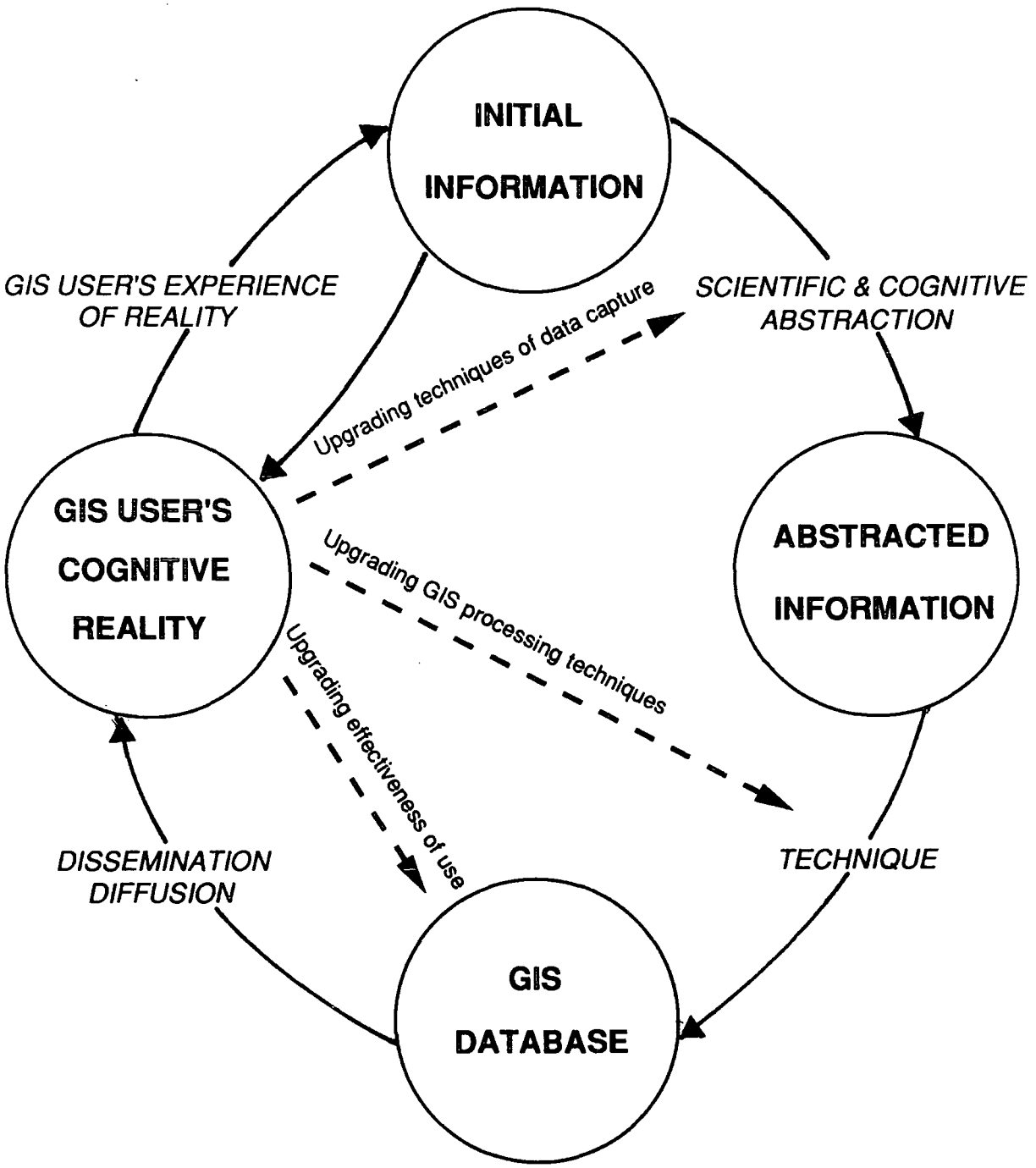
2.2 The topological data structure as used by GIMMS



Source: GIMMS Manual Version 5.0

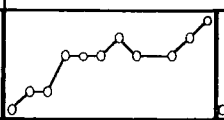
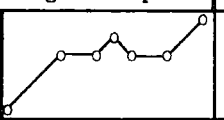
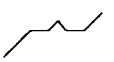
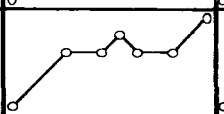
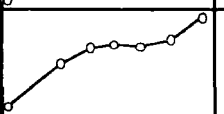
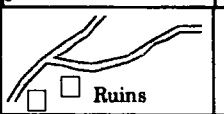
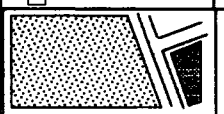
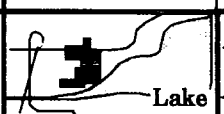
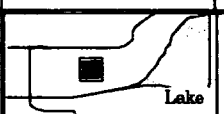
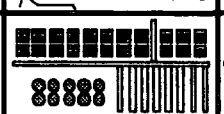
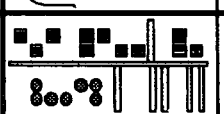
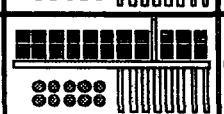
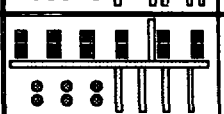
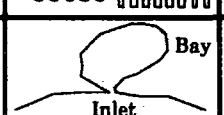
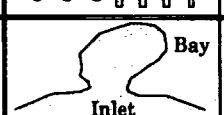
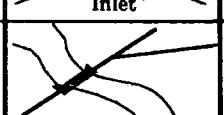
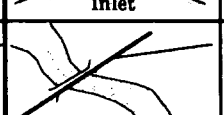
Figures for Chapter 3

3.1. A model of error transmission in a GIS.



Adapted from Blakemore and Harley 1980

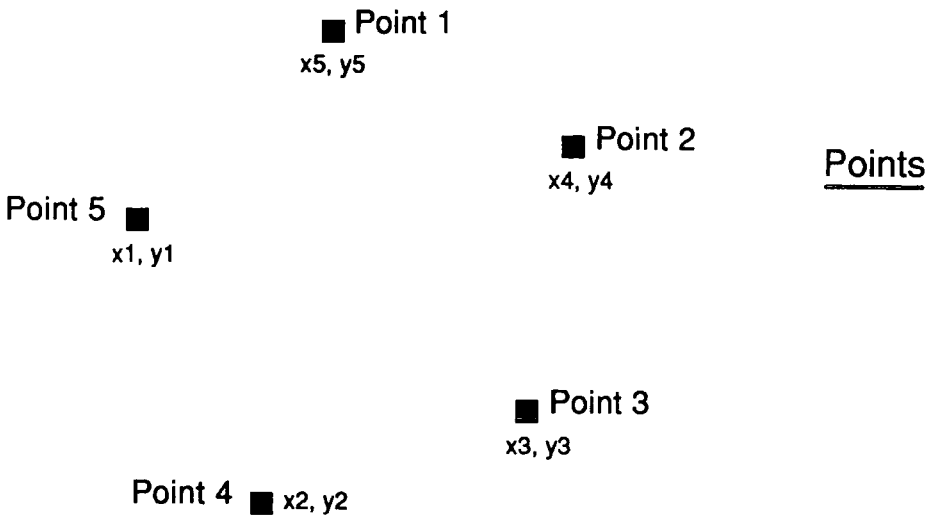
3. 2. Common methods of cartographic generalisation.

Spatial and Attribute Transformations (Generalization Operators)	Representation in the Original Map	Representation in the Generalized Map	
	At Scale of the Original Map		At 50% Scale
Simplification			
Smoothing			
Aggregation			
Amalgamation			
Merge			
Collapse			
Refinement			
Typification			
Exaggeration			
Enhancement			
Displacement			
Classification	1,2,3,4,5,6,7,8,9,10,11,12, 13,14,15,16,17,18,19,20	1-5, 6-10, 11-15, 16-20	Not Applicable

Shea and McMaster 1989

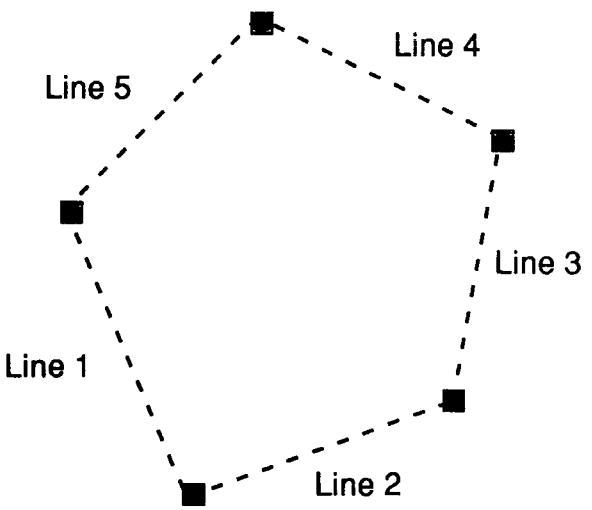
Figures for Chapter 4

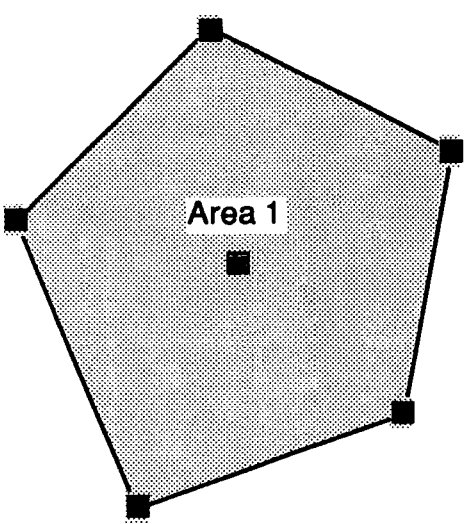
4.1 Components of digitising



Lines

Made up of points PLUS information. For example; note the position of x, y keep the pen down and move to the position x2,y2; keep the pen down and move to position x3, y3





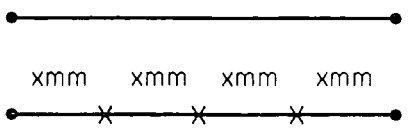
Areas

Made up of points and lines plus information to store lines 1-5 as a boundary to area 1

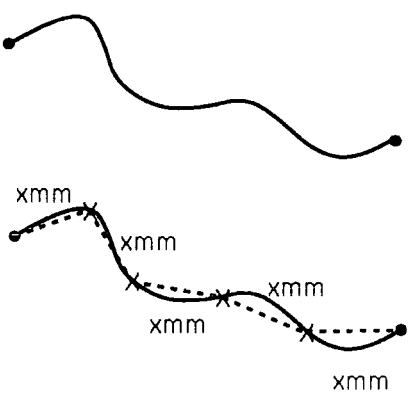
4.2 Stream Digitising

Case a: Linear stream mode

Straight line segment

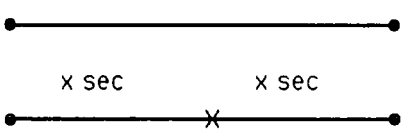


Curved line segment

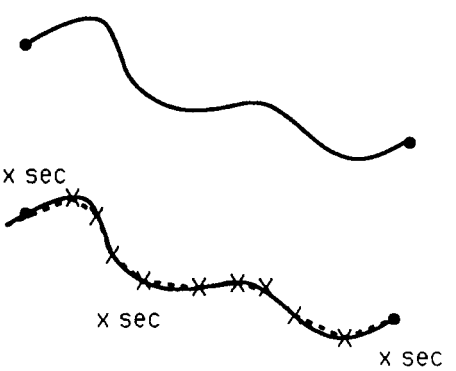


Case b: Temporal stream mode

Straight line segment



Curved line segment



4.3 Costs and benefits of three data capture methodologies

	Intelligence	Technology	Software	Labour	cost
Scanners	low	high	high	low	high
Line followers	medium	medium	medium	medium	medium / low
Manual Digitisers	high	low	low	high	medium / low

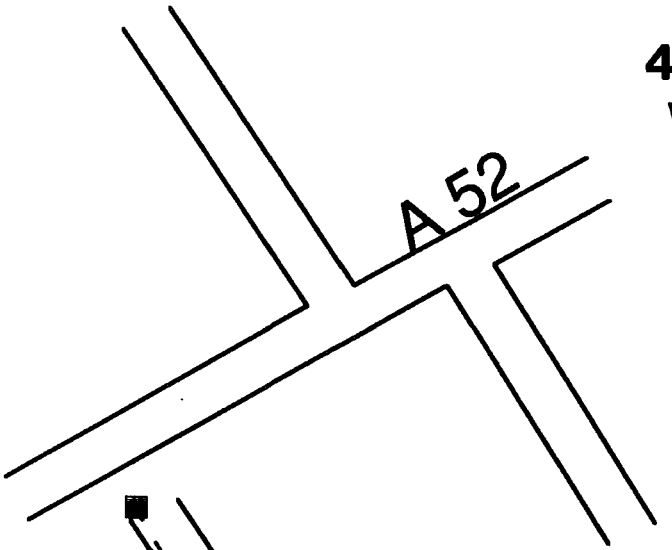
	Source material preparation	Time responsiveness	Speed of actual process	Volume advantages
Scanners	high	low	high	high
Line followers	high	low	high	high
Manual Digitisers	medium	high	low	medium / low

	Physical error distribution
Scanners	predictable
Line followers	somewhat predictable
Manual Digitisers	unpredictable

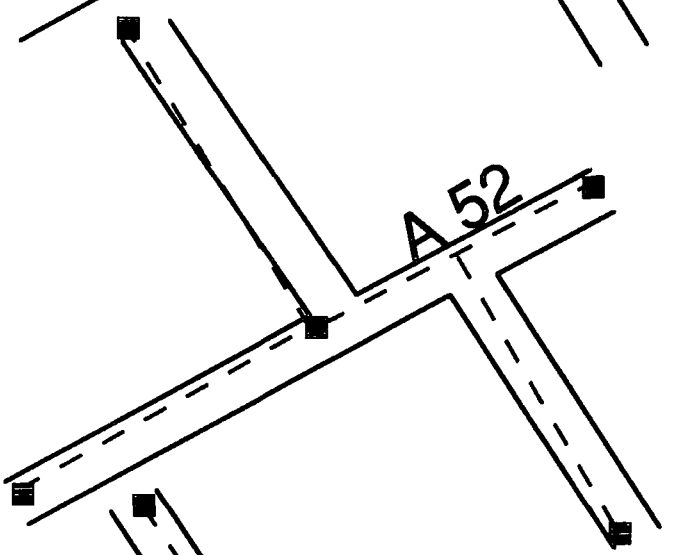
4.4 Error components of the table digitising process

Error component	Causes	Consequence
Technology	Design of soft/hardware	Resolution
Eyesight	Individual	Features unreliably represented
Hand movement	Individual	Features unreliably represented
Operational training skills	Agency	Undesirable accuracy practices
Perception and 'intelligence'	Individual/ agency?	Heavily generalised lines/too many points stored/line poorly caricatured
Scale	Source material	Features poorly represented
Copying and translation	Individual/ agency	Features unreliably represented
Physical components of map sheet	Source material	Mismatches at map joins
Mistakes	Individual	Features unreliably represented

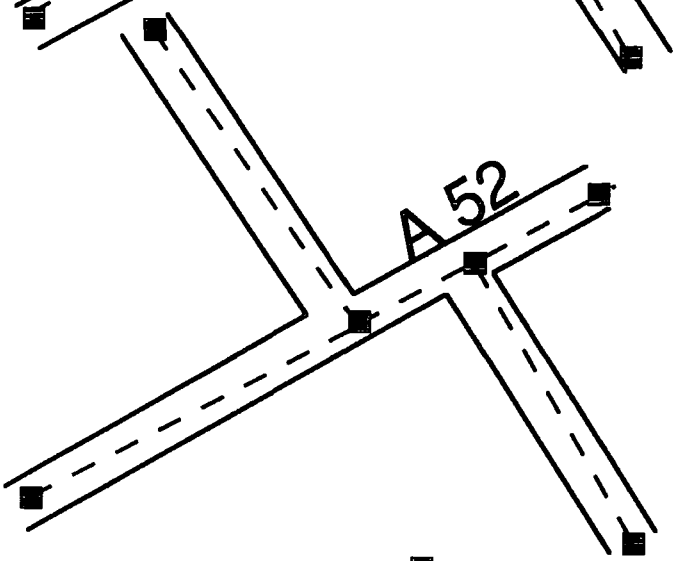
4.5 Problems of line following when digitising



Linear feature to be digitised

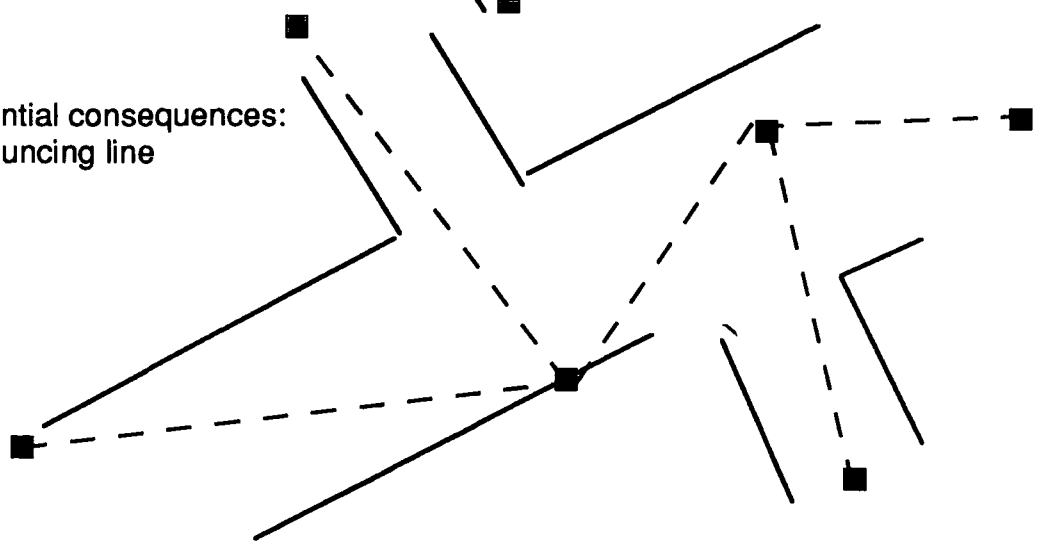


Option a: using one boundary as a guide

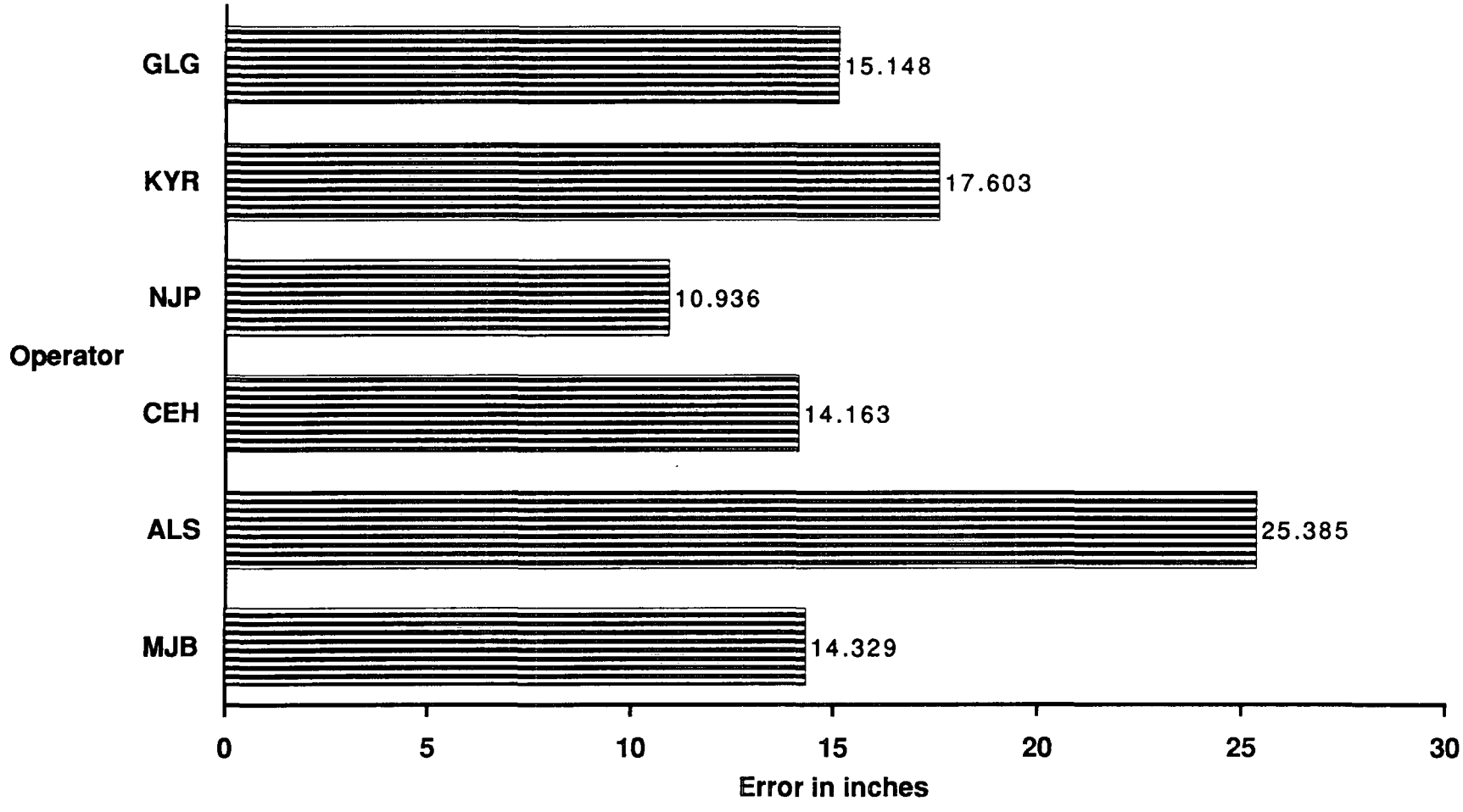


Option b: attempting a centreline policy

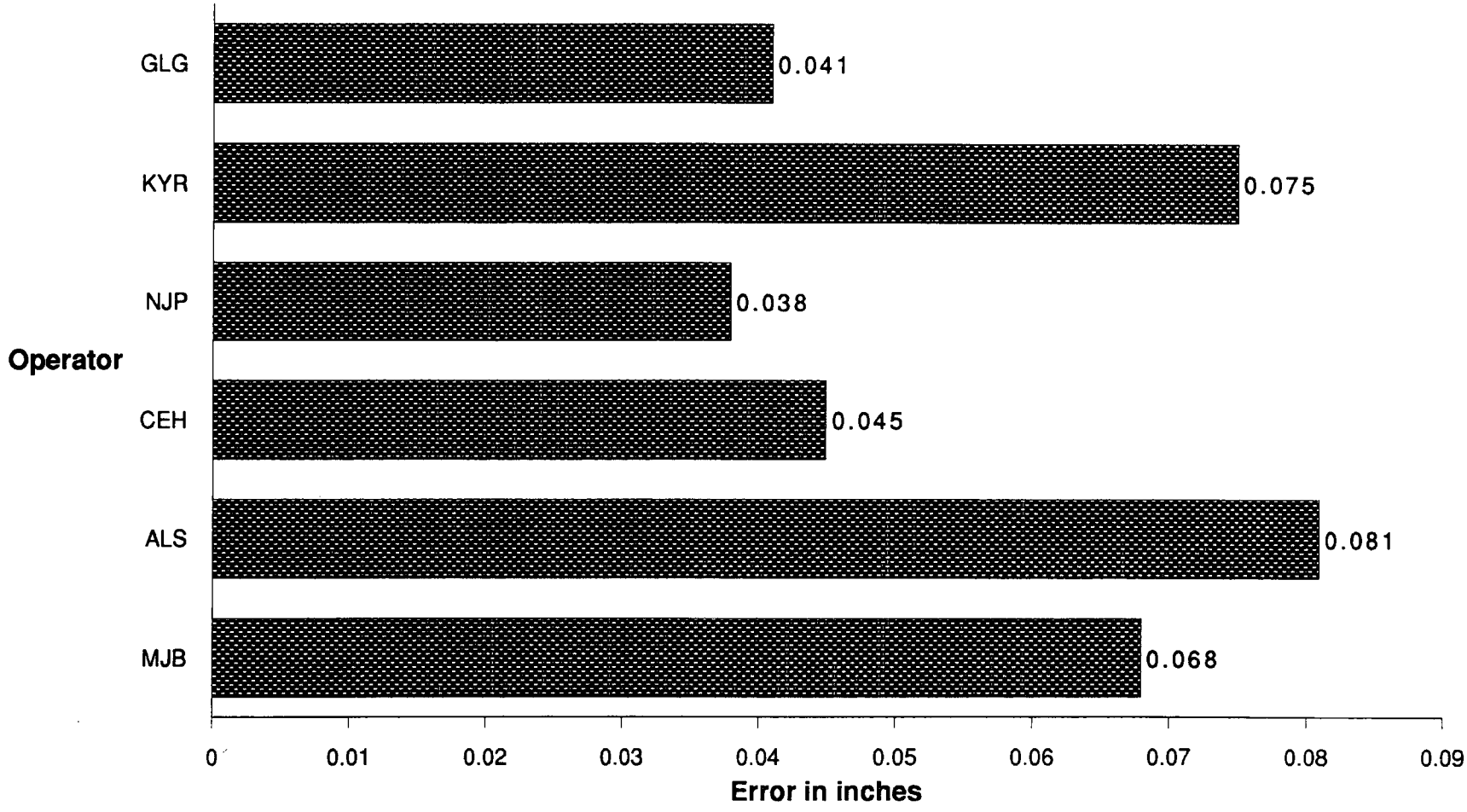
Potential consequences:
A bouncing line



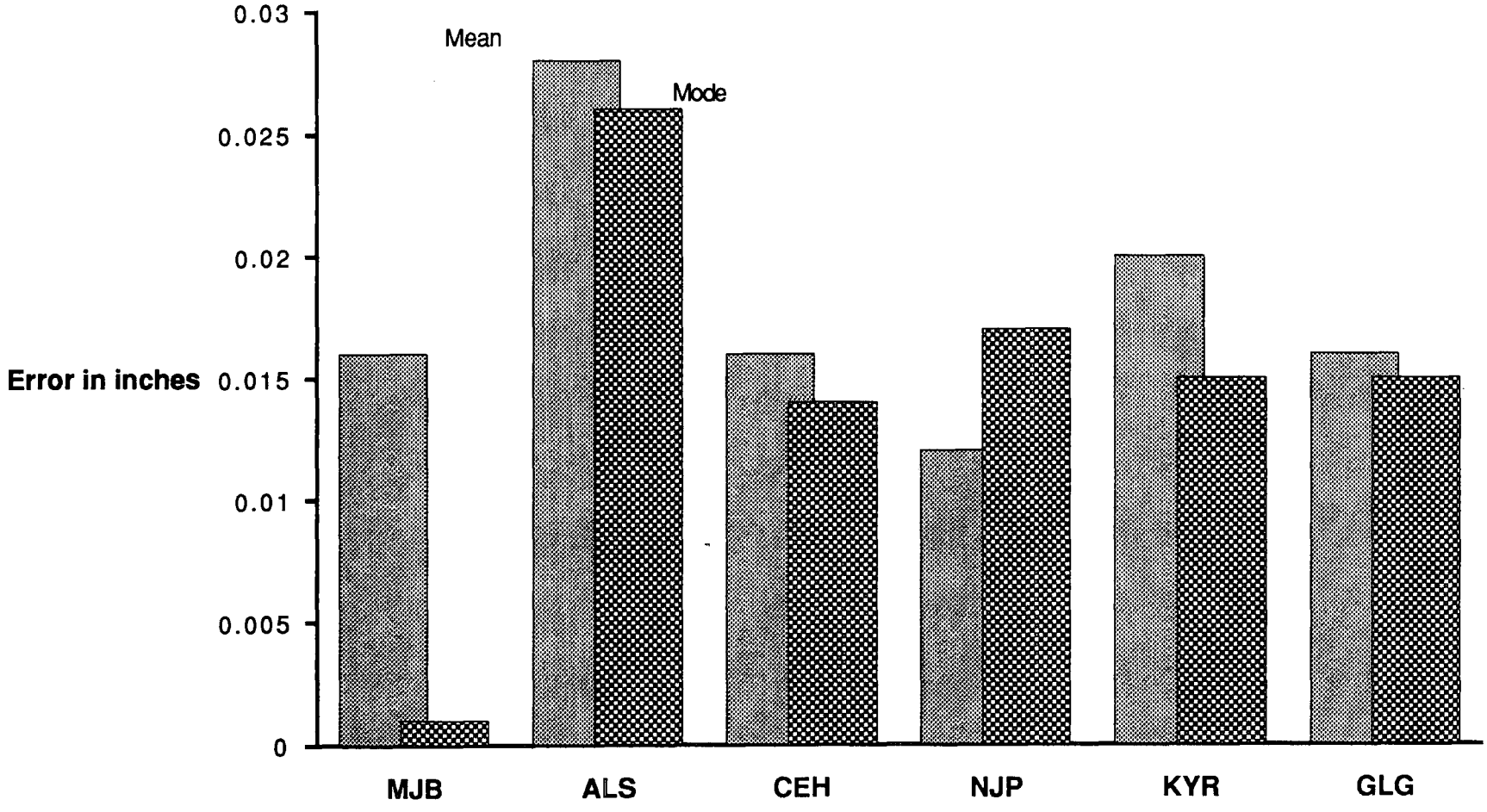
4.6 Variations in the sum of error for the entire grid



4.7 Variations in the total deviation for individual points



4.8 Mean and mode values for operators in the grid experiment



4.9

A frequency distribution of the total error encountered by operator MJB

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  4.00 OCCURRENCES

   77      .00067  ******************
  111      .00400  ***************************
  133      .00733  *********************************
   98      .01067  ************************
  106      .01400  **************************
   74      .01733  ******************
   68      .02067  *****************
   56      .02400  **************
   44      .02733  ***********
   38      .03067  **********
   27      .03400  *******
   19      .03733  *****
   17      .04067  ****
    9      .04400  **
    9      .04733  **
    2      .05067  *
    6      .05400  **
    3      .05733  *
    1      .06067
    1      .06400
    1      .06733
                I_________I_________I_________I_________I_________I
                0         40        80        120       160       200
                          HISTOGRAM FREQUENCY
```

4.10

A frequency distribution of the total error encountered by operator ALS

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  2.00 OCCURRENCES

   32        .000  ****************
   70        .004  ***********************************
   60        .008  ******************************
   60        .012  ******************************
   65        .016  *********************************
   59        .020  *****************************
   68        .024  **********************************
   96        .028  ************************************************
   65        .032  *********************************
   67        .036  **********************************
   62        .040  *******************************
   57        .044  ****************************
   43        .048  **********************
   19        .052  **********
   32        .056  ****************
   16        .060  ********
   11        .064  ******
   10        .068  *****
    3        .072  **
    4        .076  **
    2        .080  *
                I____|____I____|____I____|____I____|____I____|____I
                0         20        40        60        80        100
                          HISTOGRAM FREQUENCY
```

4.11

A frequency distribution of the total error encountered by operator CEH

```
COUNT    MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  2.00 OCCURRENCES

    0     -.0025
   21      .0000  ***********
   54      .0025  ***************************
   62      .0050  *******************************
   72      .0075  ************************************
   88      .0100  ********************************************
   80      .0125  ****************************************
   90      .0150  *********************************************
   90      .0175  *********************************************
   99      .0200  **************************************************
   91      .0225  *********************************************
   63      .0250  ********************************
   34      .0275  *****************
   28      .0300  **************
   17      .0325  *********
    4      .0350  **
    2      .0375  *
    1      .0400  *
    2      .0425  *
    2      .0450  *
    0      .0475
                  I____|____I____|____I____|____I____|____I____|____I
                  0         20        40        60        80       100
                            HISTOGRAM FREQUENCY
```

4.12

A frequency distribution of the total error encountered by operator NJP

```
COUNT    MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  2.00 OCCURRENCES

    0      -.001
   74       .001  *************************************
   90       .003  *********************************************
   73       .005  ************************************
   89       .007  *********************************************
   72       .009  ************************************
   79       .011  ****************************************
   87       .013  *********************************************
   67       .015  ***********************************
   71       .017  ************************************
   53       .019  ****************************
   32       .021  ****************
   35       .023  ******************
   21       .025  ***********
   17       .027  *********
   12       .029  ******
   12       .031  ******
    7       .033  ****
    3       .035  **
    5       .037  ***
    1       .039  *
                  I____|____I____|____I____|____I____|____I____|____I
                  0         20        40        60        80       100
                            HISTOGRAM FREQUENCY
```

4.13

A frequency distribution of the total error encountered by operator KYR

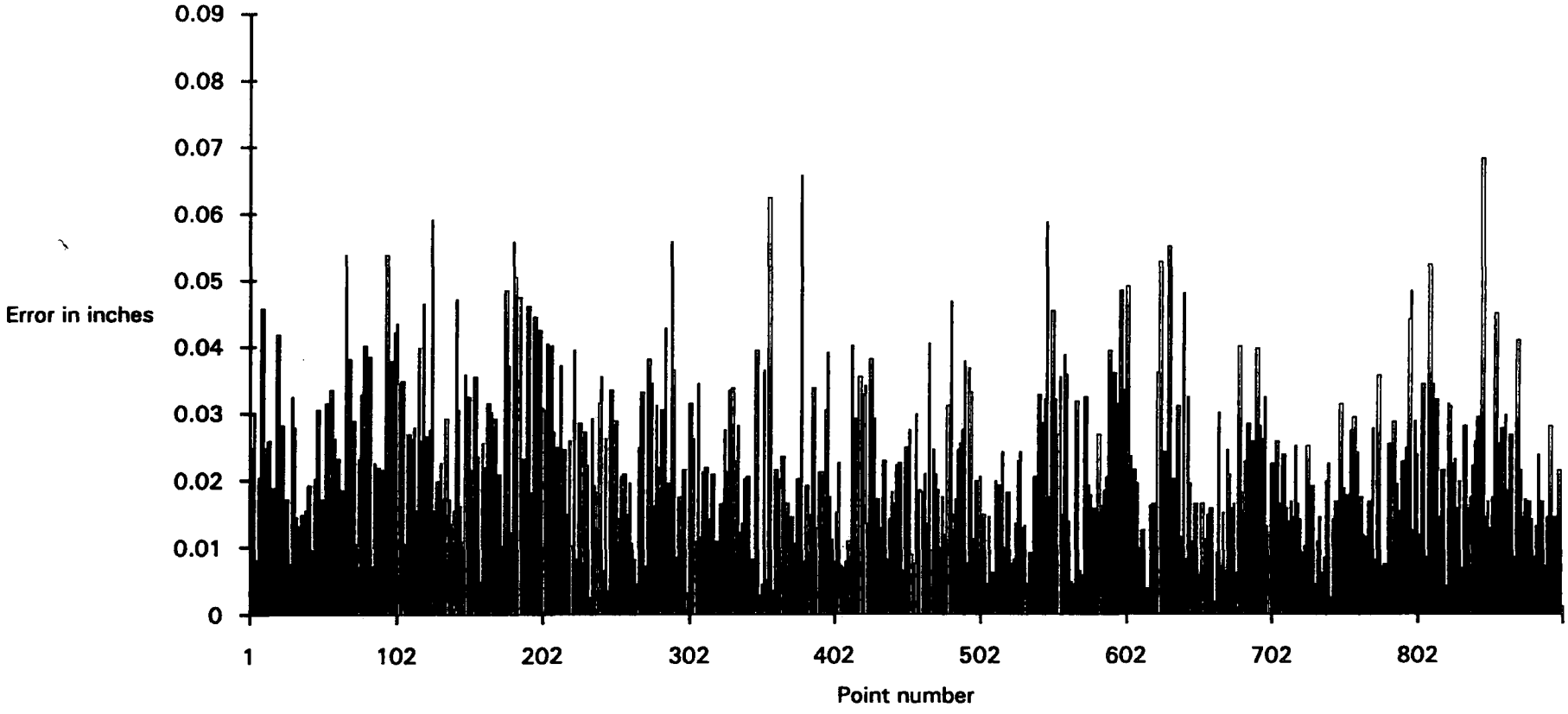
```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  4.00 OCCURRENCES

    0      -.002
  127       .002  **********************************
  112       .006  ******************************
   93       .010  *************************
   93       .014  *************************
   86       .018  ************************
   76       .022  ********************
   74       .026  ********************
   60       .030  ****************
   54       .034  **************
   35       .038  *********
   36       .042  *********
   19       .046  *****
   16       .050  ****
    6       .054  **
    5       .058  *
    4       .062  *
    2       .066  *
    1       .070
    1       .074
    0       .078
              I____|____I____|____I____|____I____|____I____|____I____|____I
              0         40         80         120        160        200
                          HISTOGRAM FREQUENCY
```

4.14

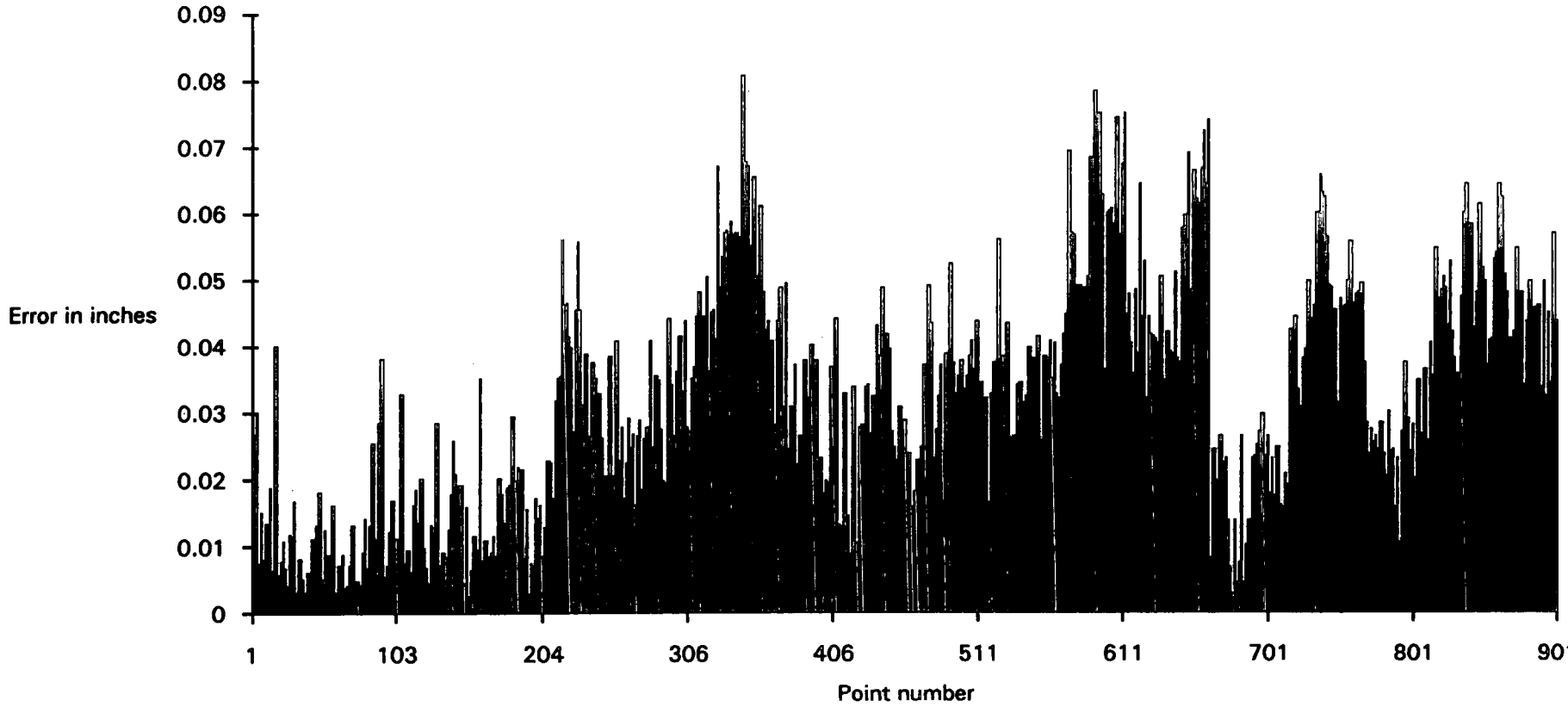
A frequency distribution of the total error encountered by operator GLG

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  2.00 OCCURRENCES

    3       .001  **
   16       .003  ********
   32       .005  ****************
   39       .007  ********************
   70       .009  ***********************************
   90       .011  *********************************************
   94       .013  **********************************************
   86       .015  *******************************************
   91       .017  **********************************************
   93       .019  **********************************************
   79       .021  ****************************************
   64       .023  ********************************
   46       .025  ************************
   32       .027  ****************
   18       .029  *********
   19       .031  **********
   10       .033  *****
    8       .035  ****
    8       .037  ****
    2       .039  *
    1       .041  *
              I____|____I____|____I____|____I____|____I____|____I____|____I
              0         20         40         60         80         100
                          HISTOGRAM FREQUENCY
```

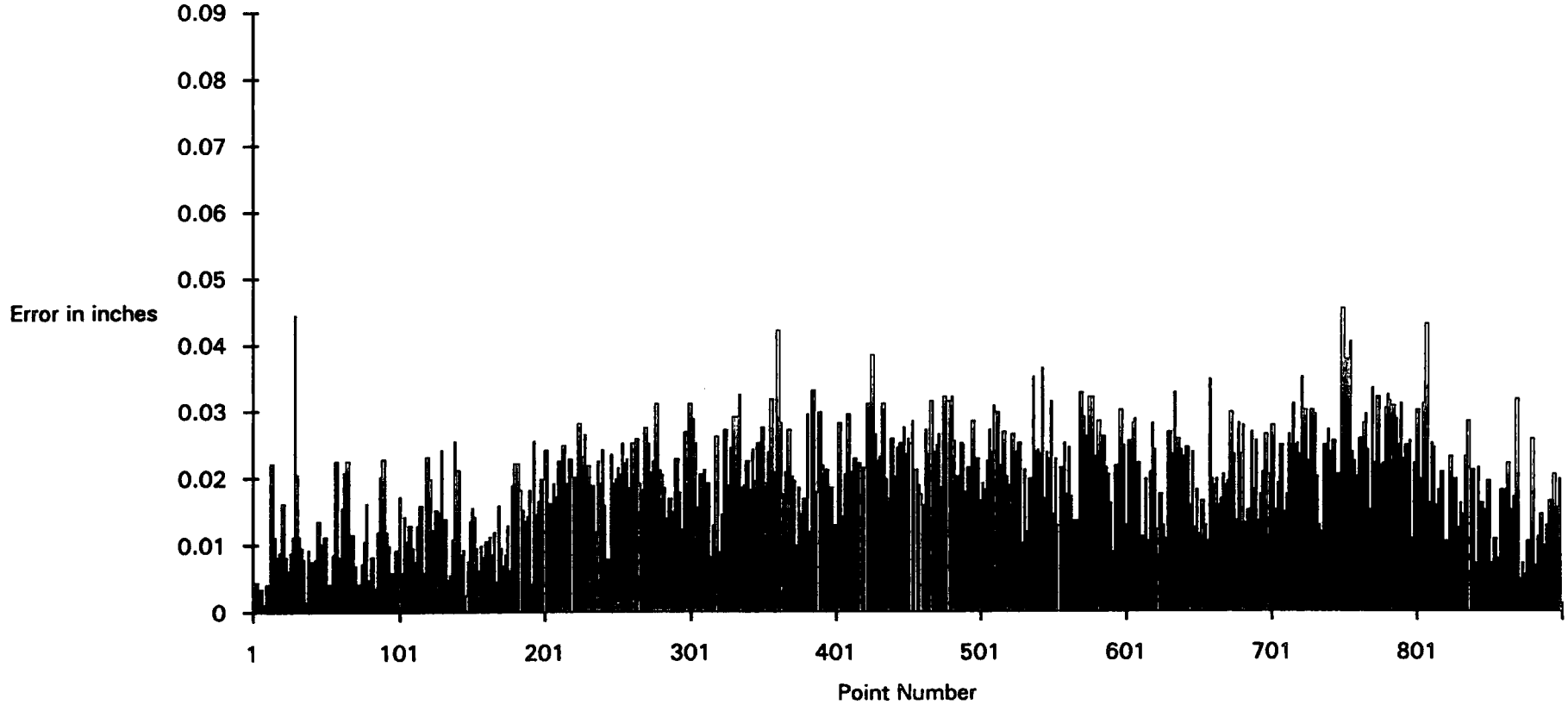
4.15 Variations in the X AND Y coordinates achieved by MJB whilst digitising



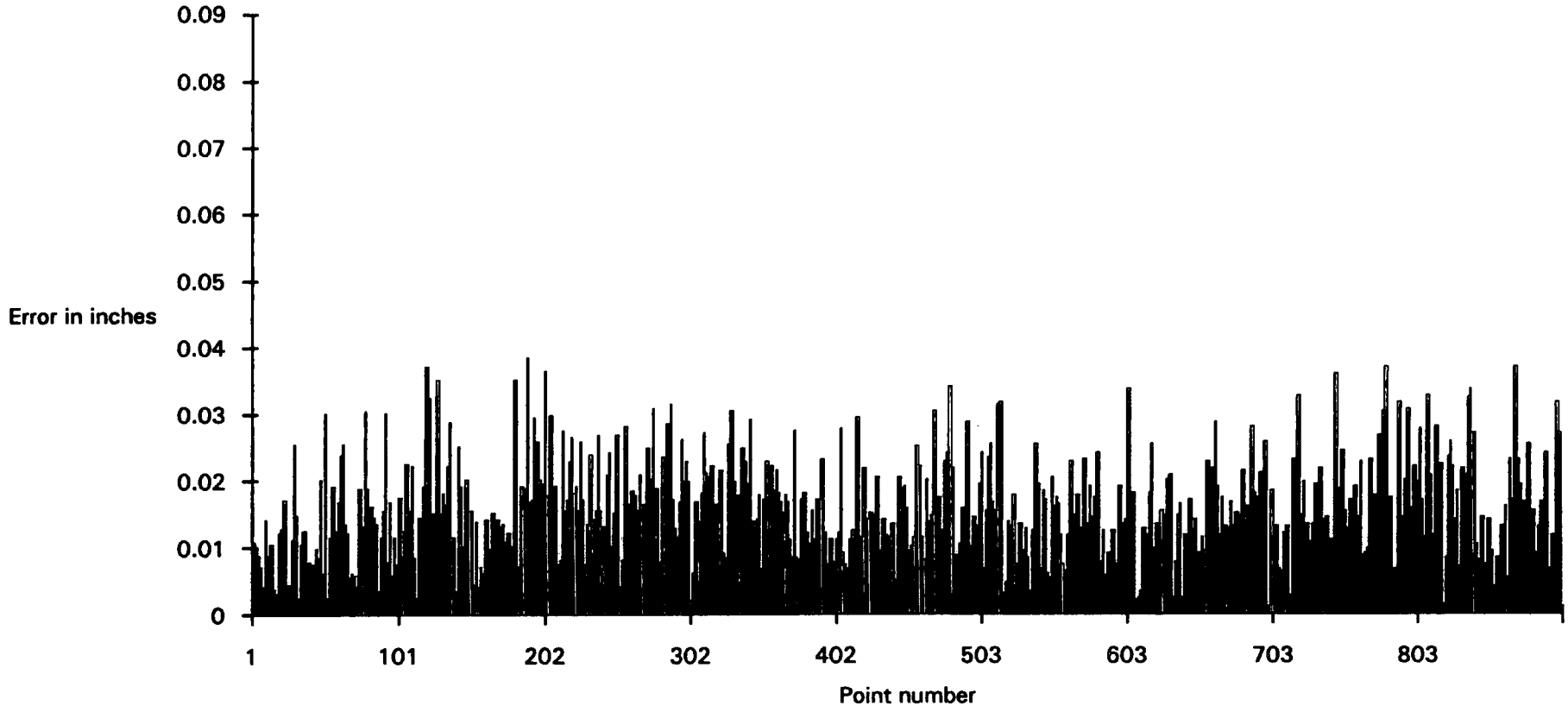
4.16 Variations in the X AND Y cooridinates achieved by ALS whilst digitising



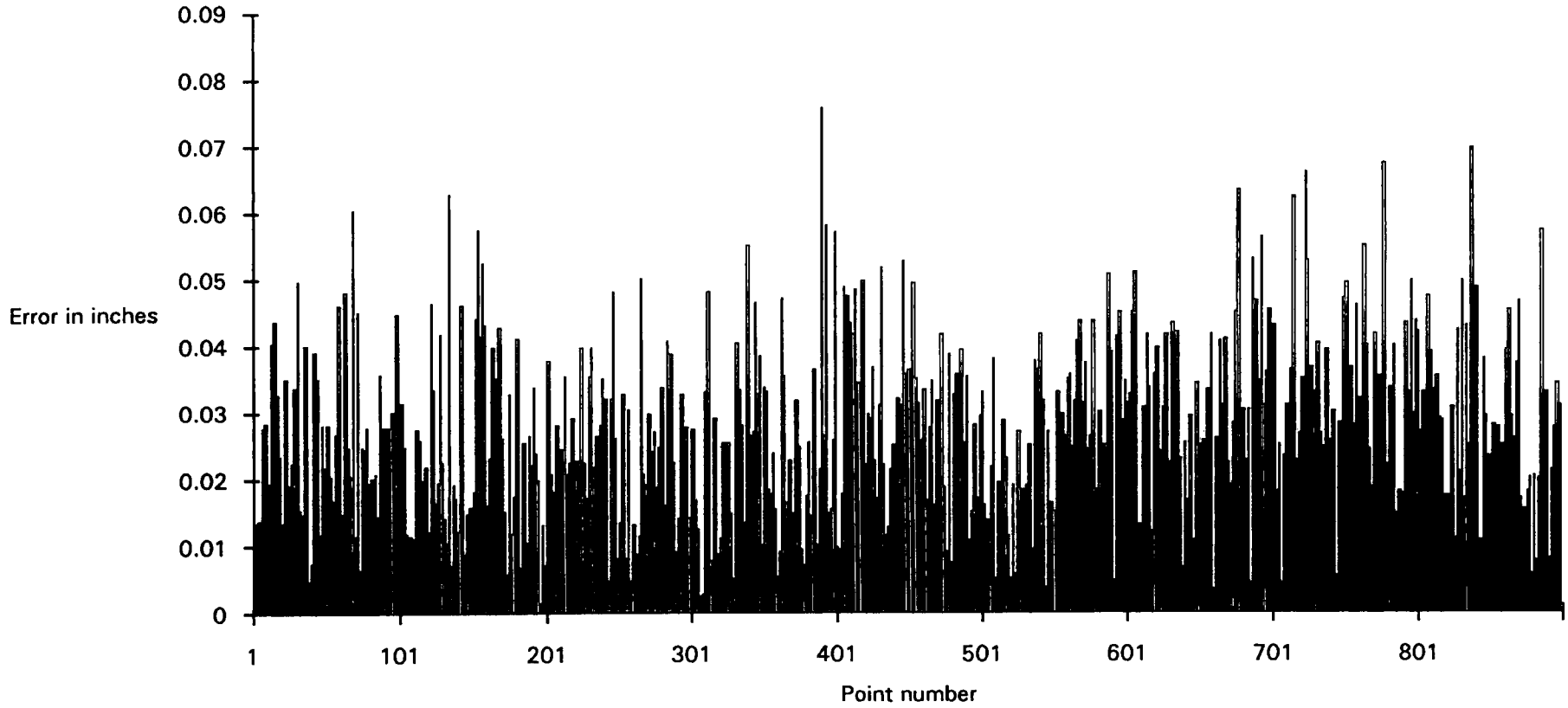
4.17 Variation in the X AND Y coordinates achieved by CEH whilst digitising.



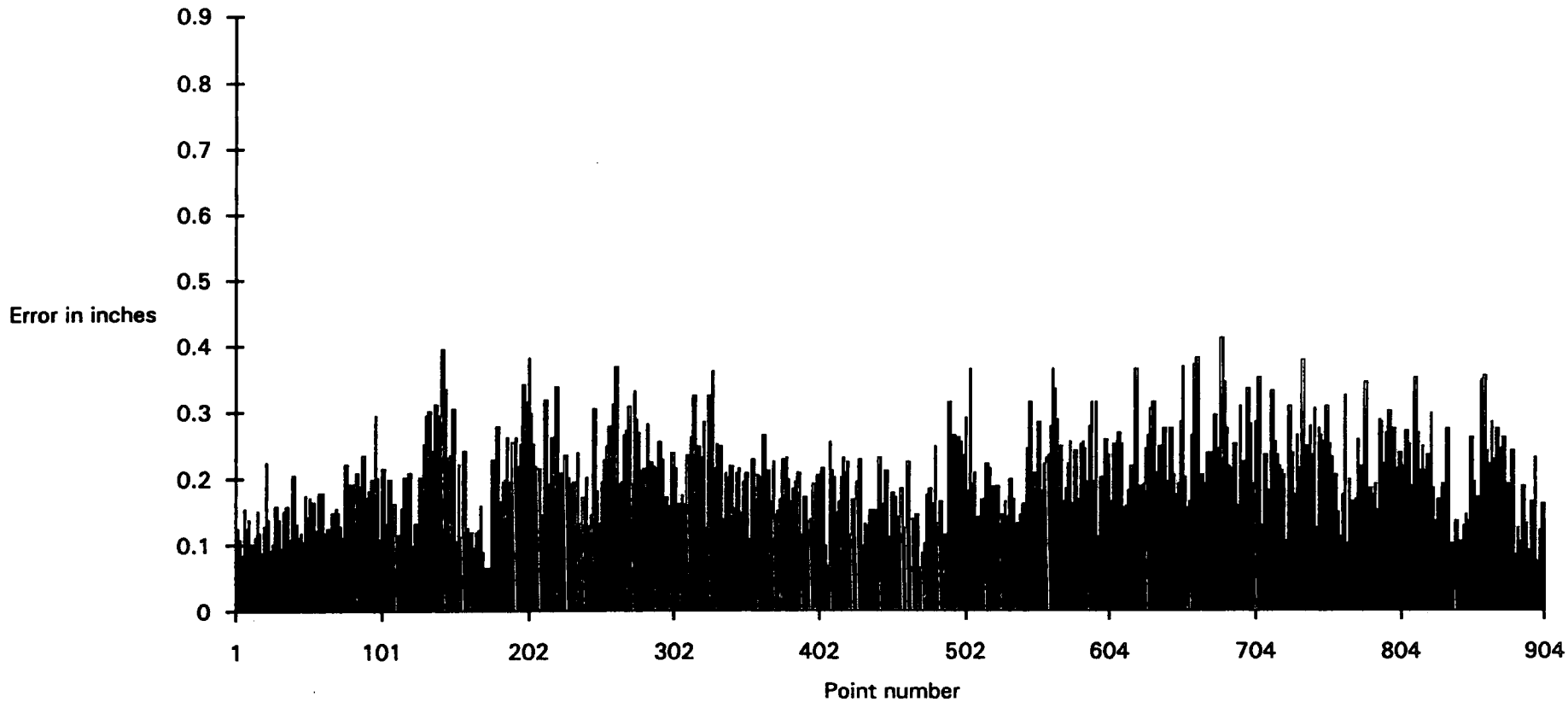
4.18 Variations in the X AND Y coordinates achieved by NJP whilst digitising



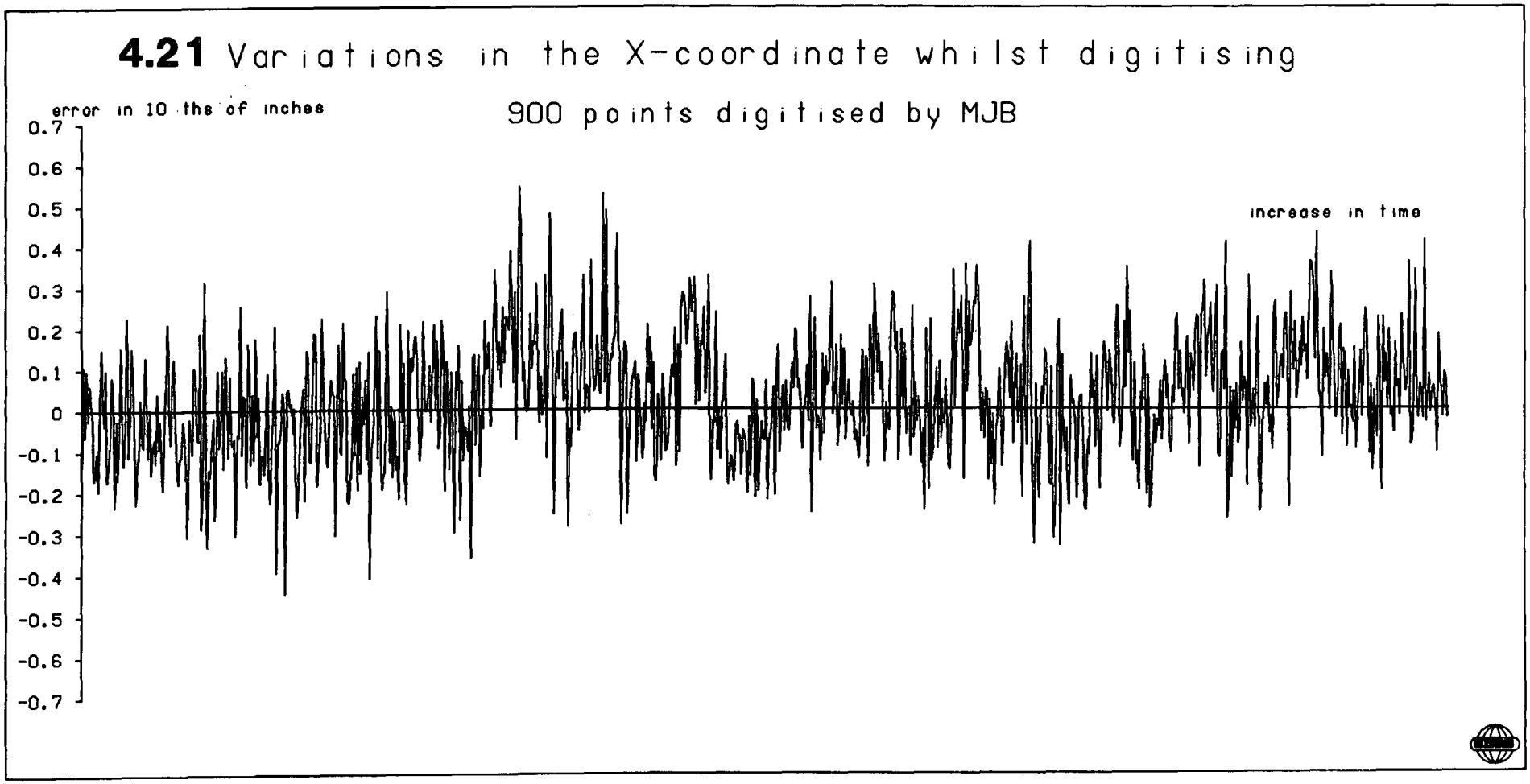
4.19 Variations in the X AND Y coordinates achieved by KYR whilst digitising



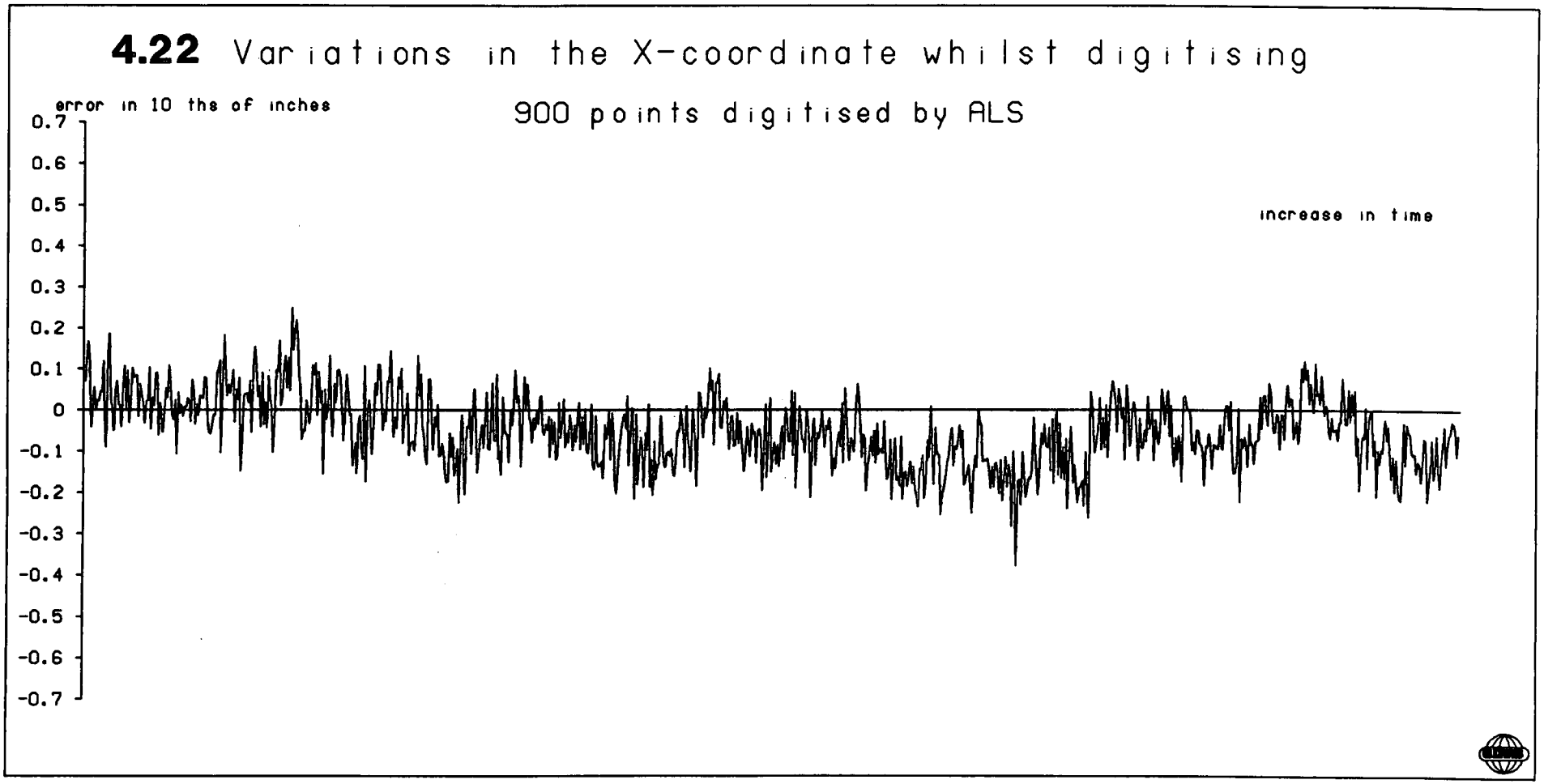
4.20 Variations in the X AND Y coordinates achieved by GLG whilst digitising



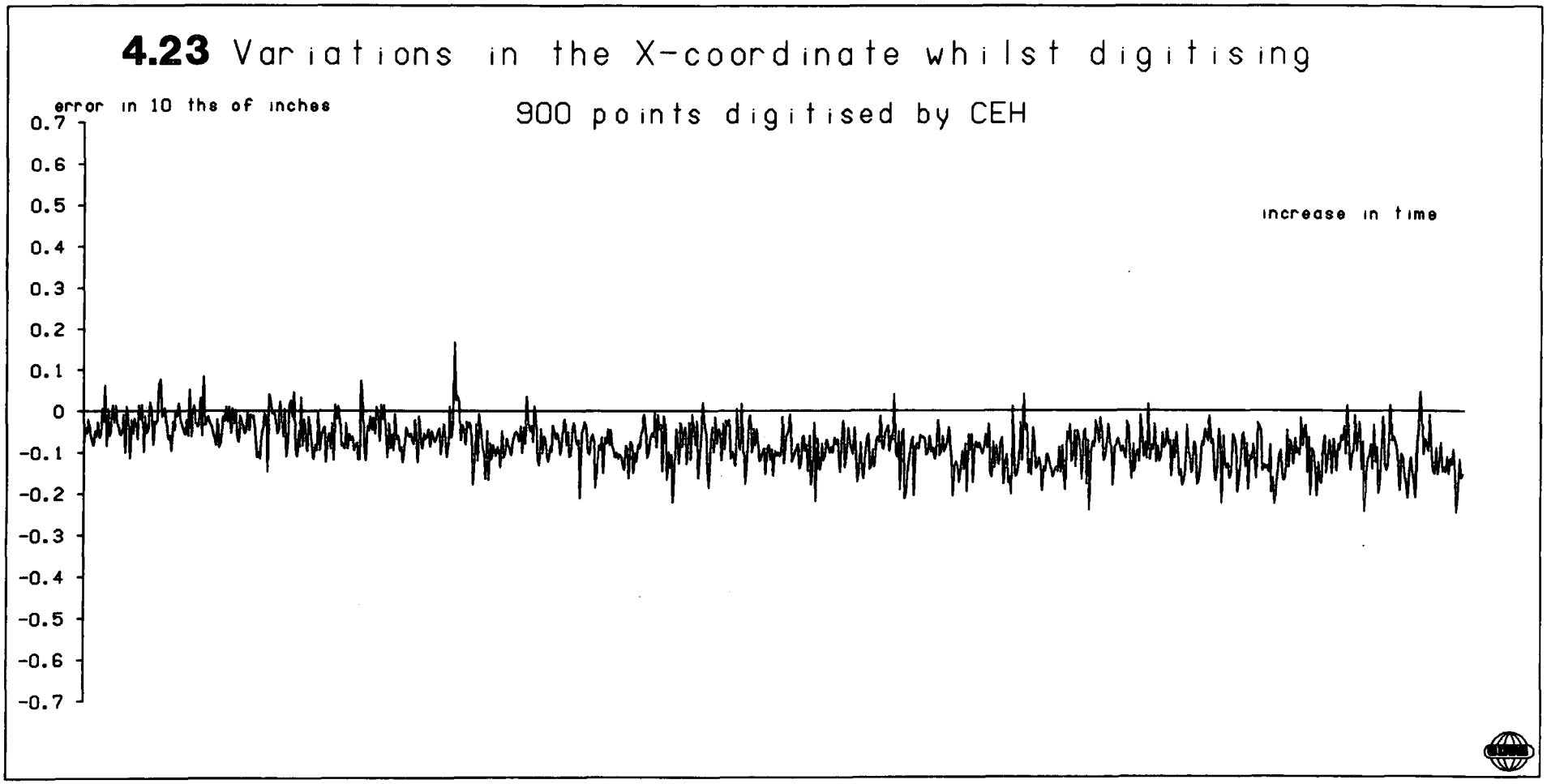
4.21 Variations in the X-coordinate whilst digitising



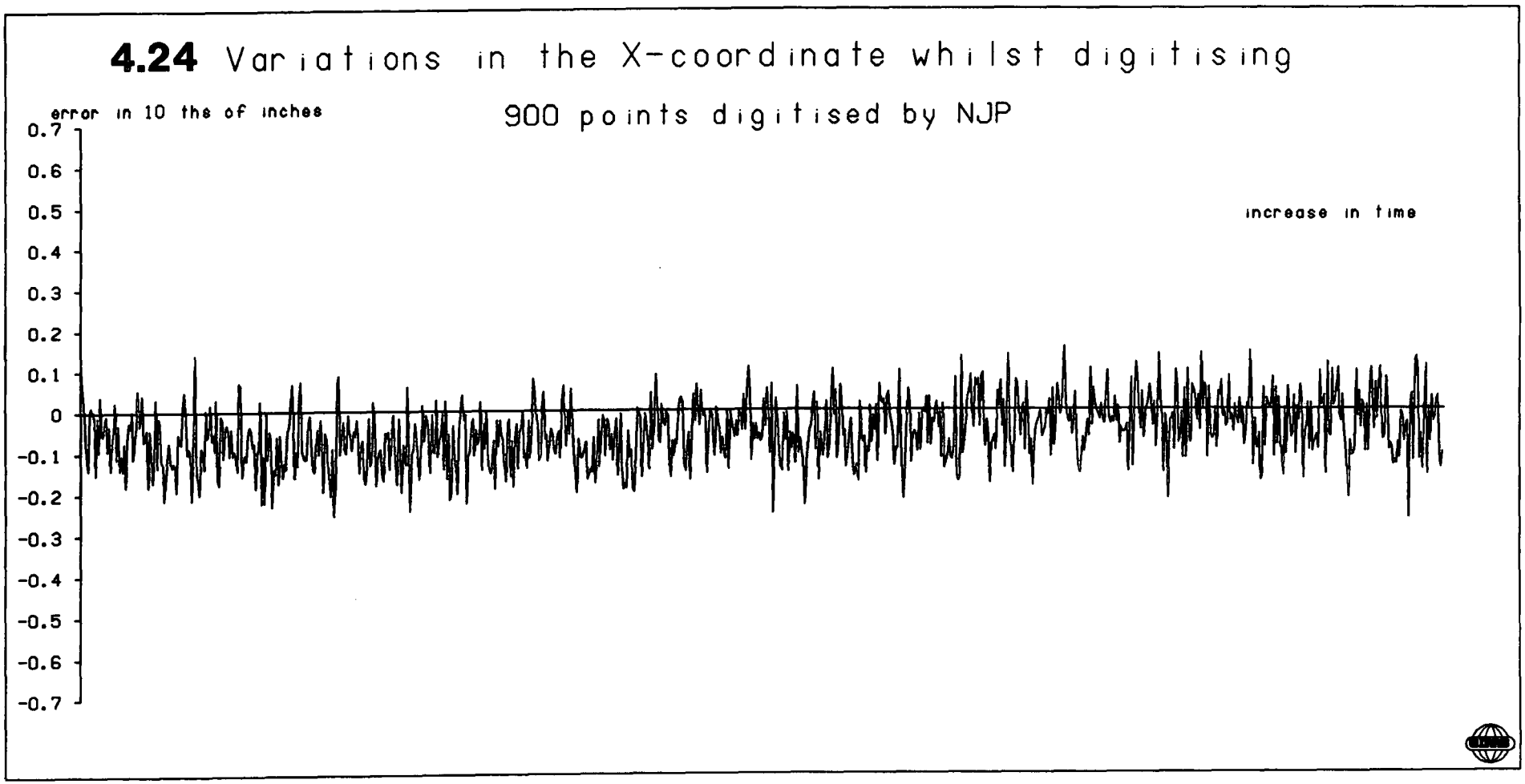
4.22 Variations in the X-coordinate whilst digitising



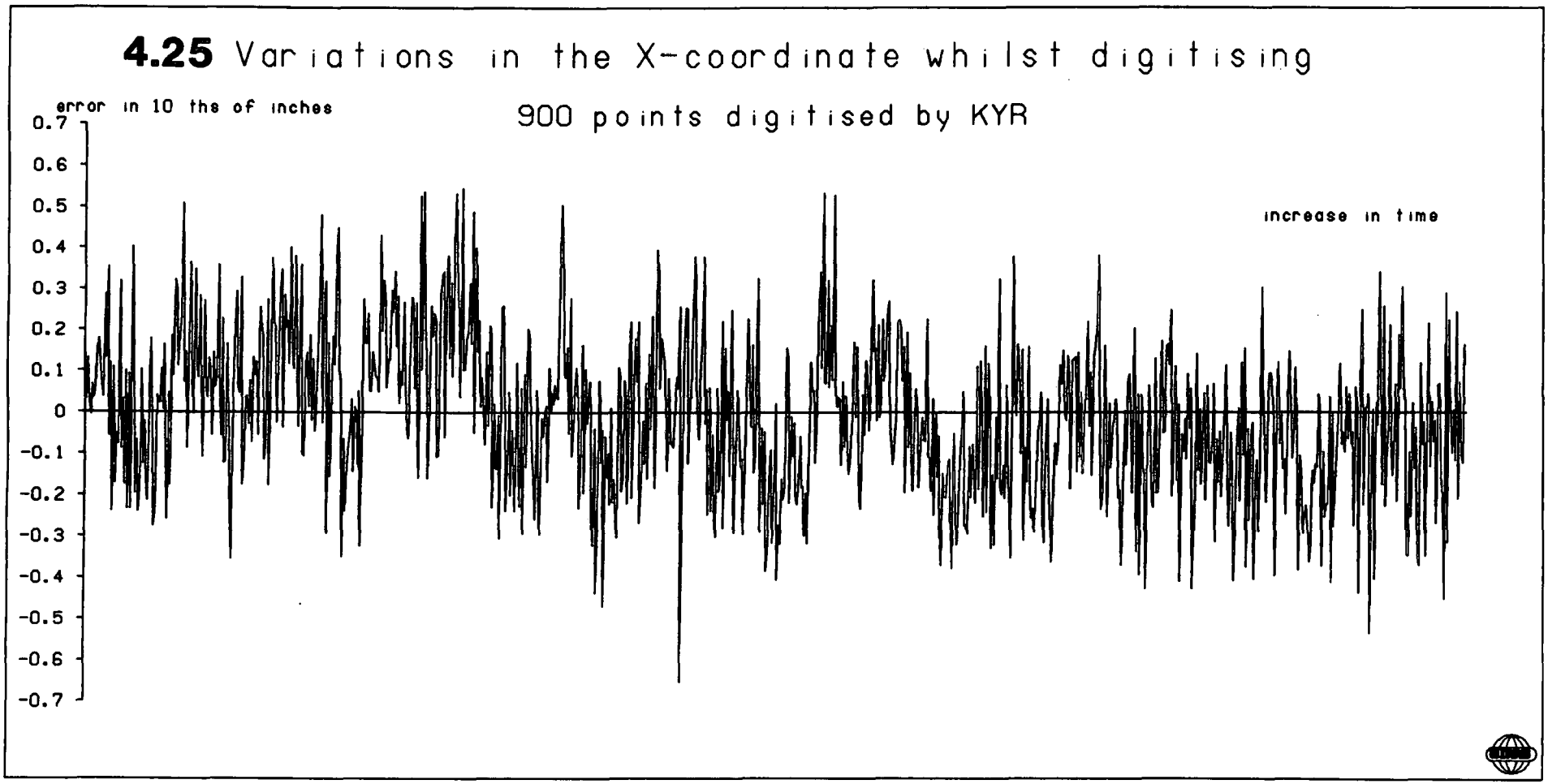
4.23 Variations in the X-coordinate whilst digitising



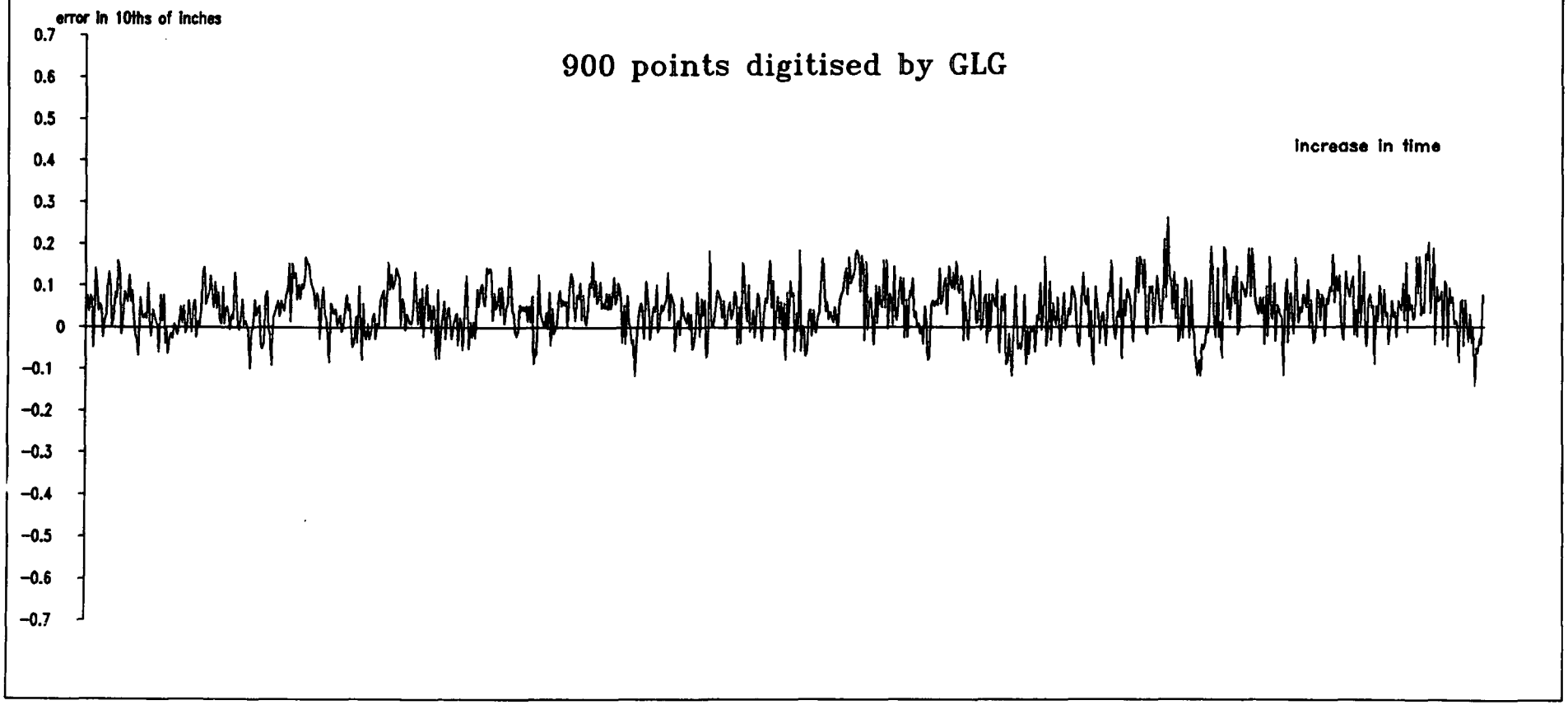
4.24 Variations in the X-coordinate whilst digitising



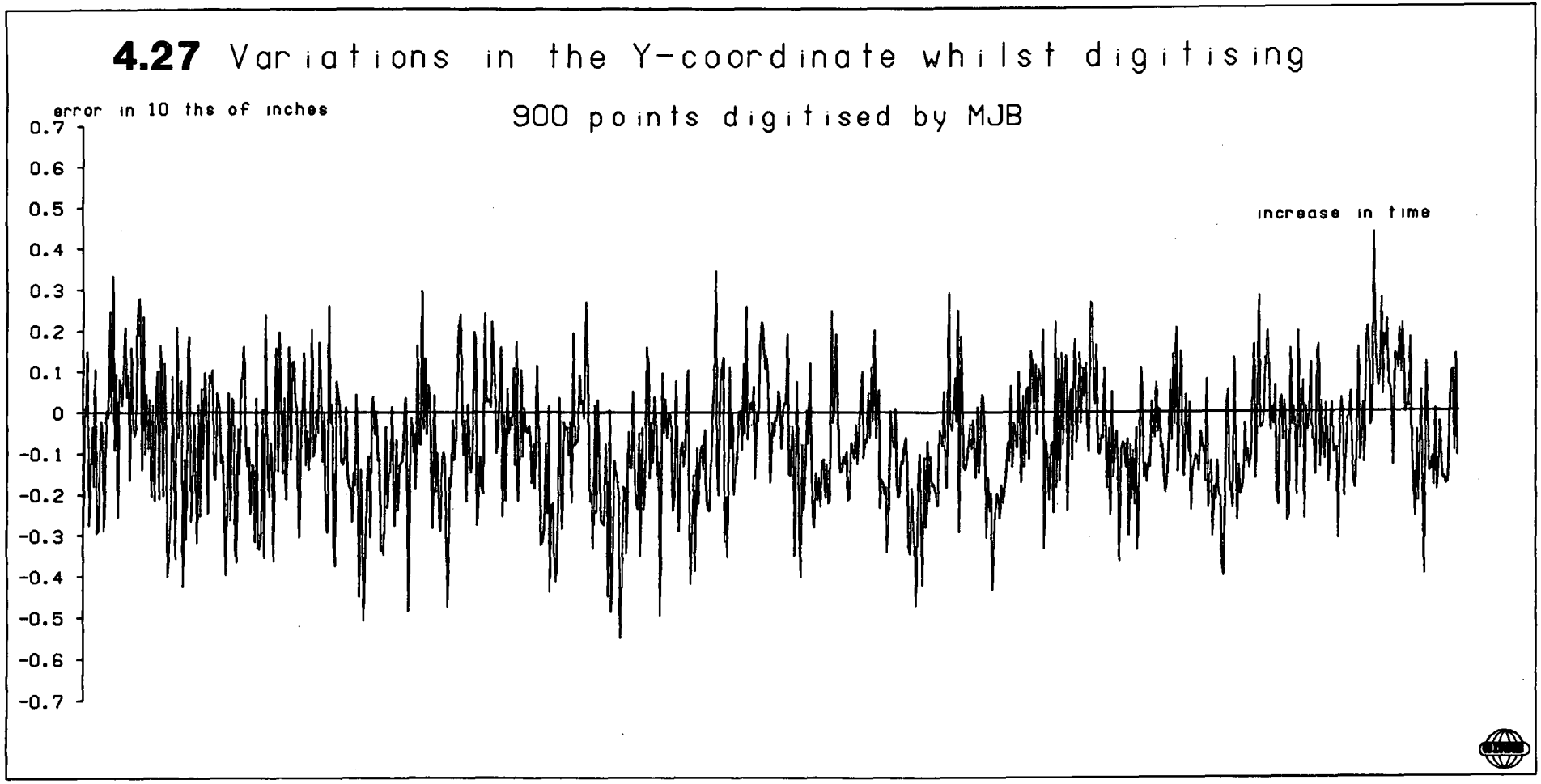
4.25 Variations in the X-coordinate whilst digitising



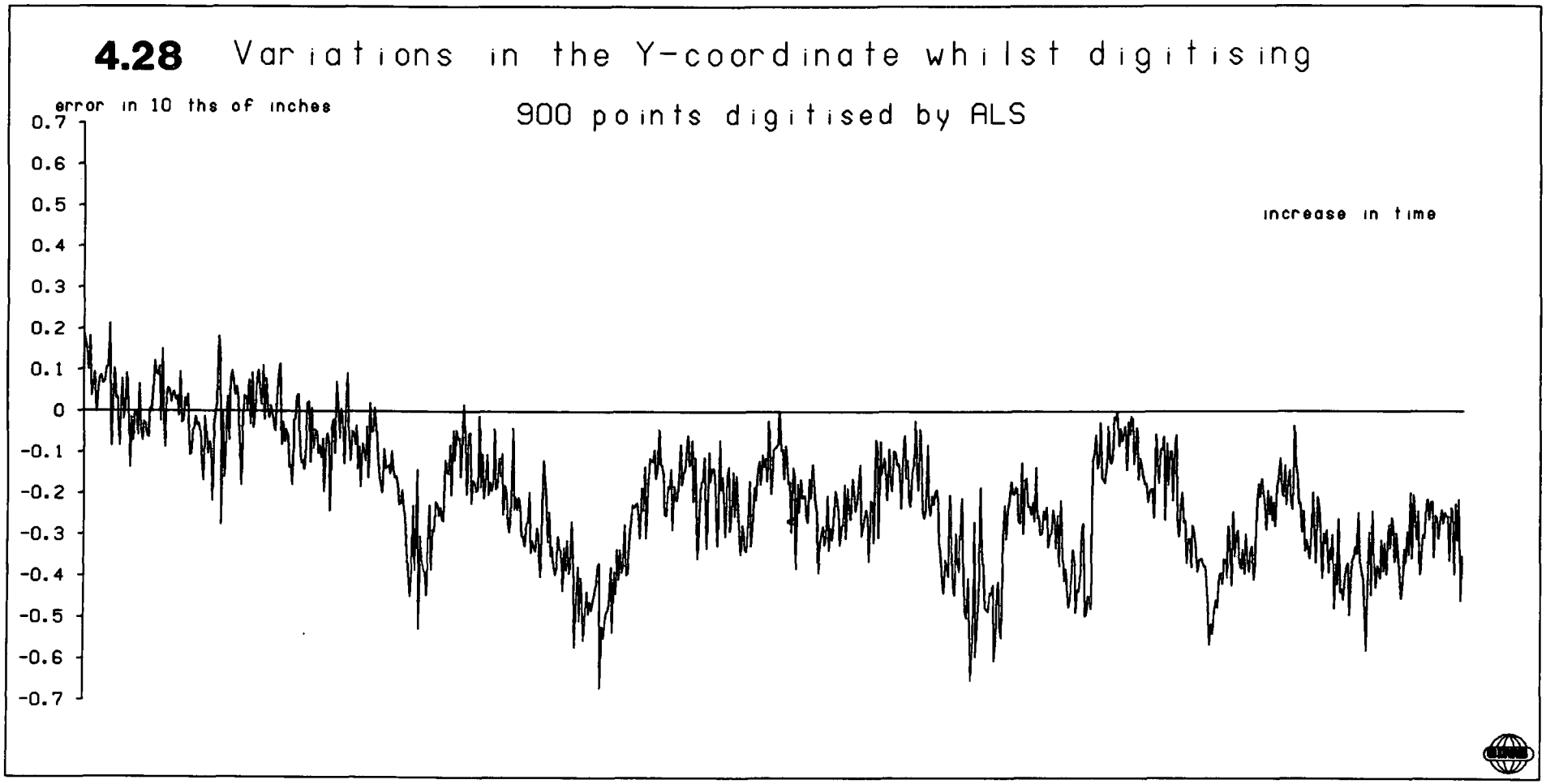
4.26 Variations in the X coordinate whilst digitising

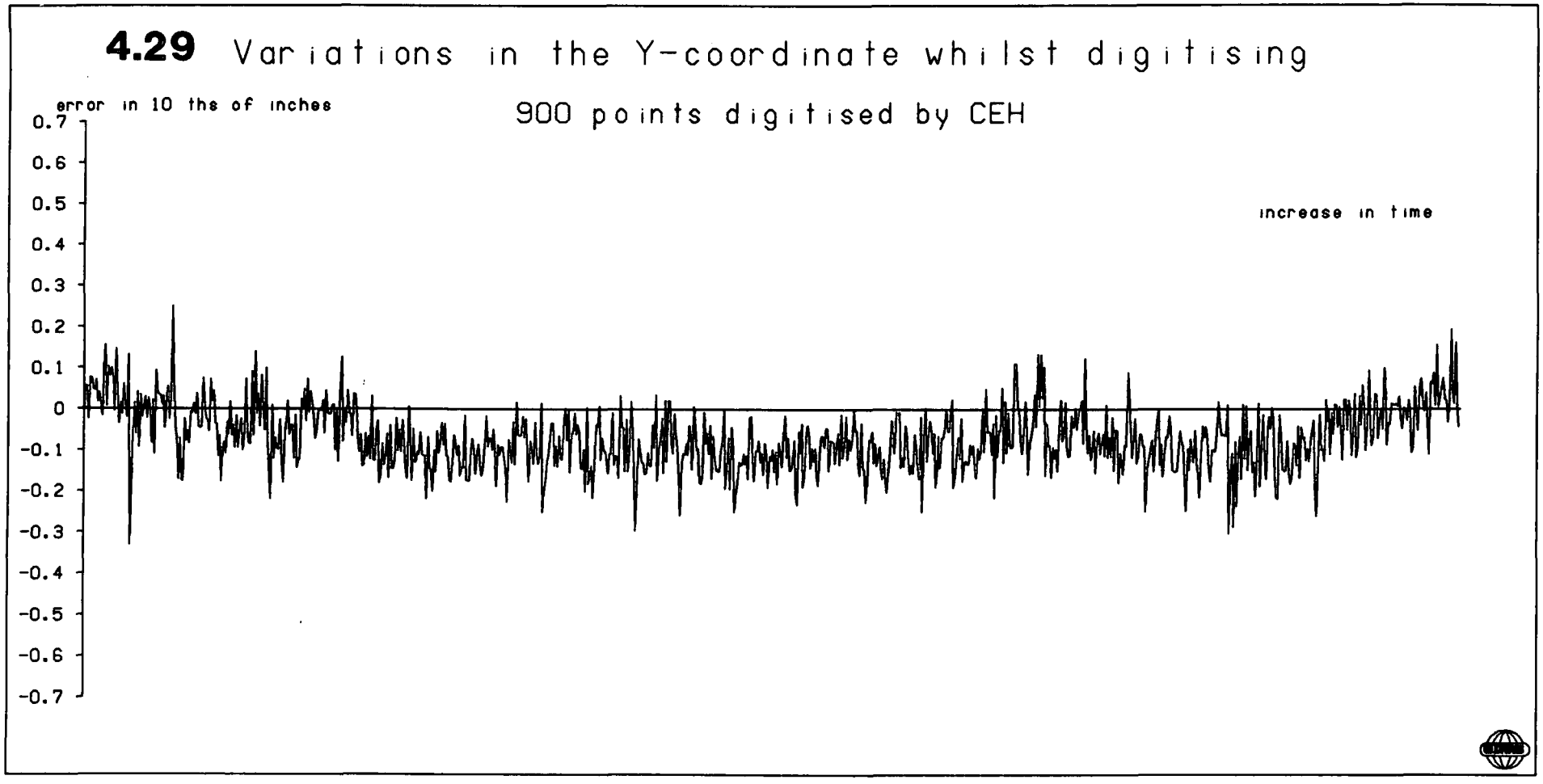


4.27 Variations in the Y-coordinate whilst digitising

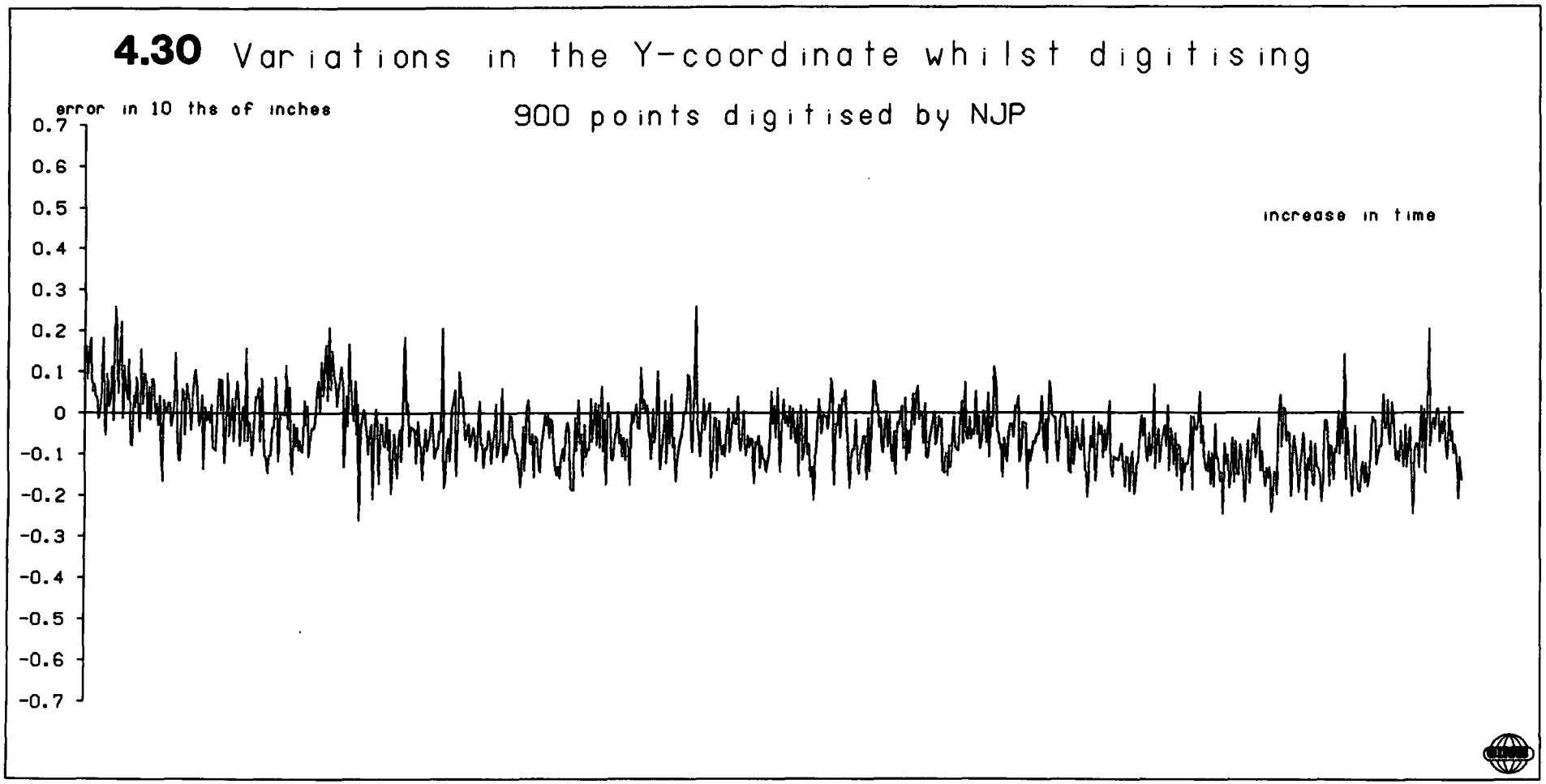


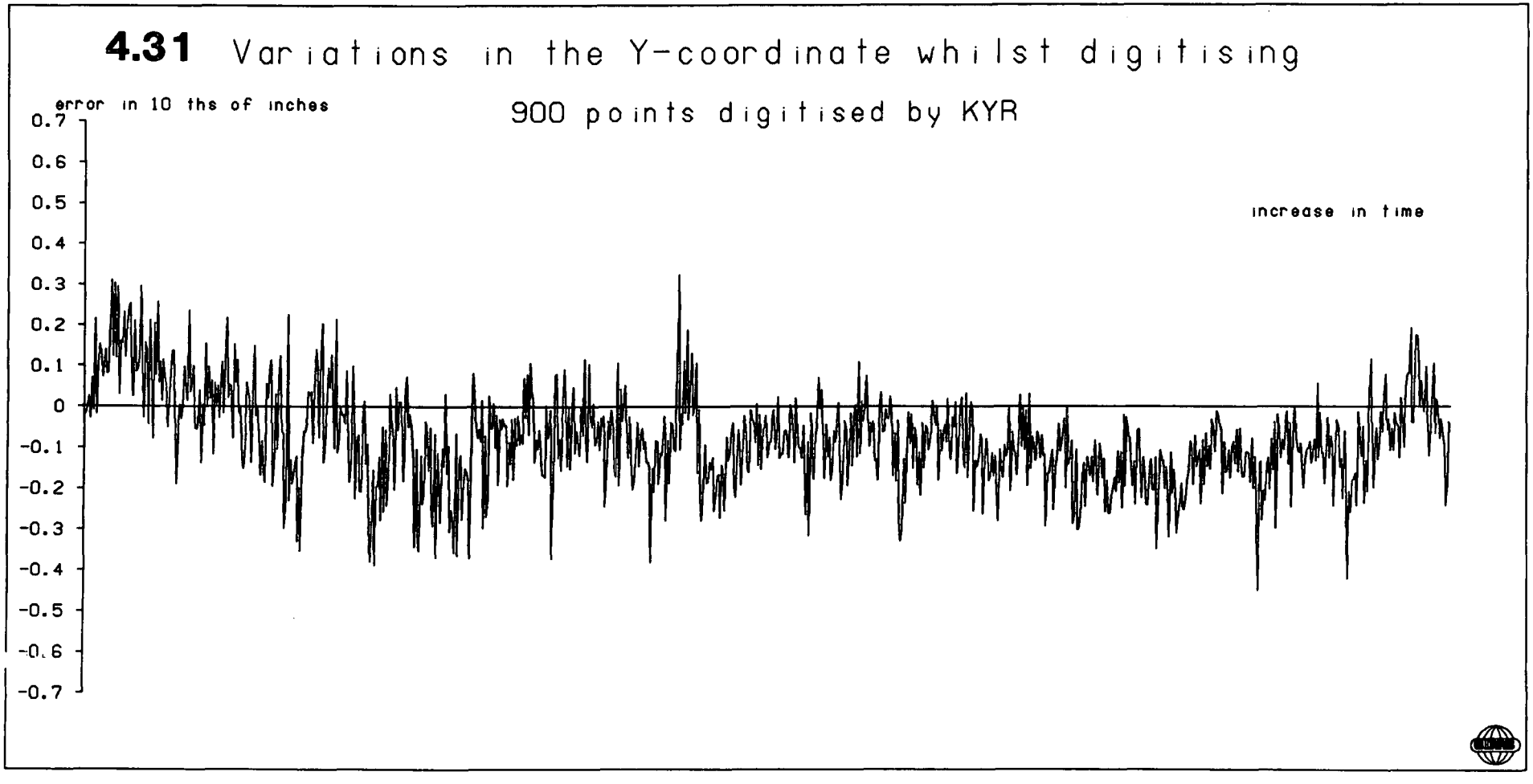
4.28 Variations in the Y-coordinate whilst digitising



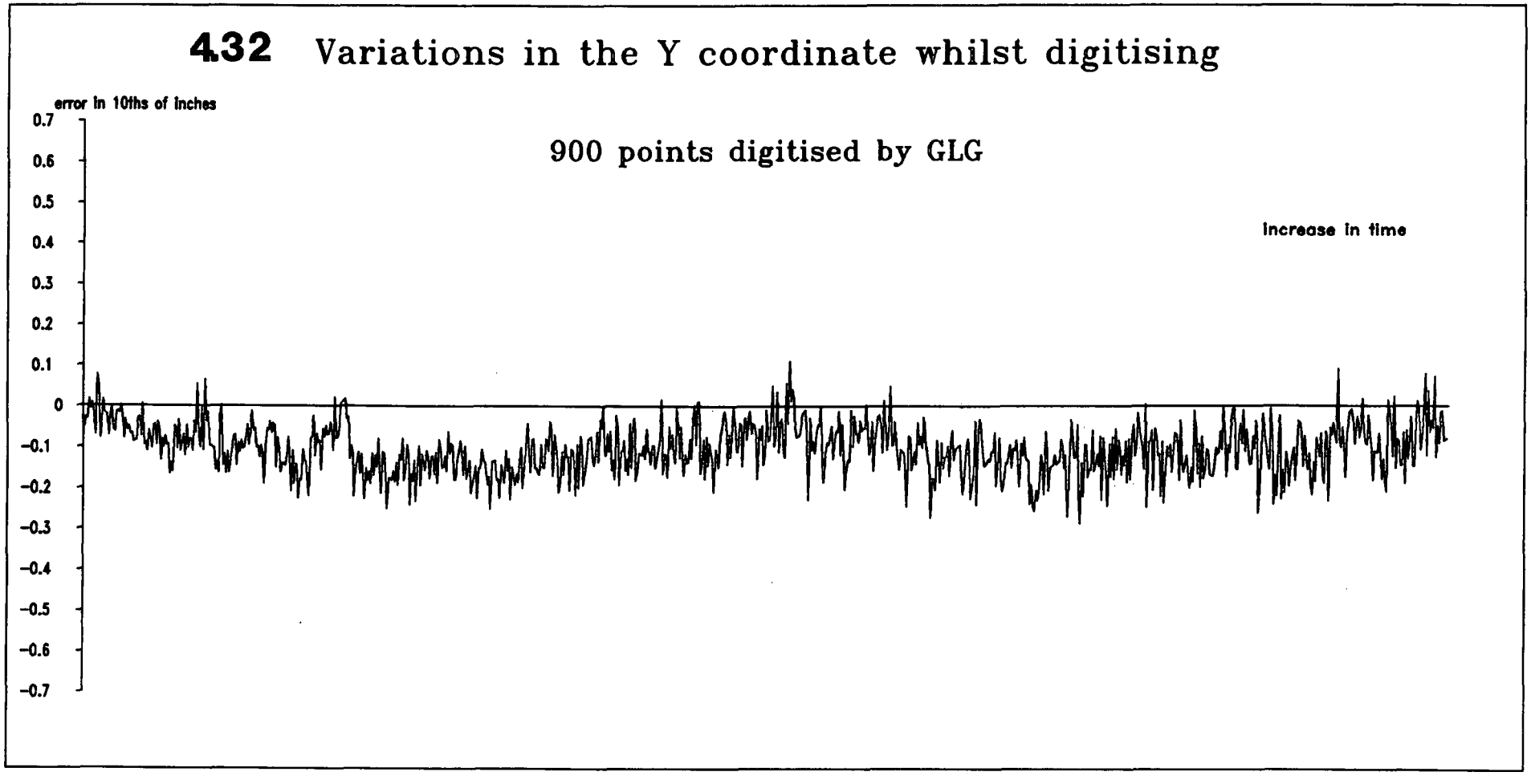

4.29 Variations in the Y-coordinate whilst digitising
900 points digitised by CEH
error in 10 ths of inches
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
increase in time

4.30 Variations in the Y-coordinate whilst digitising

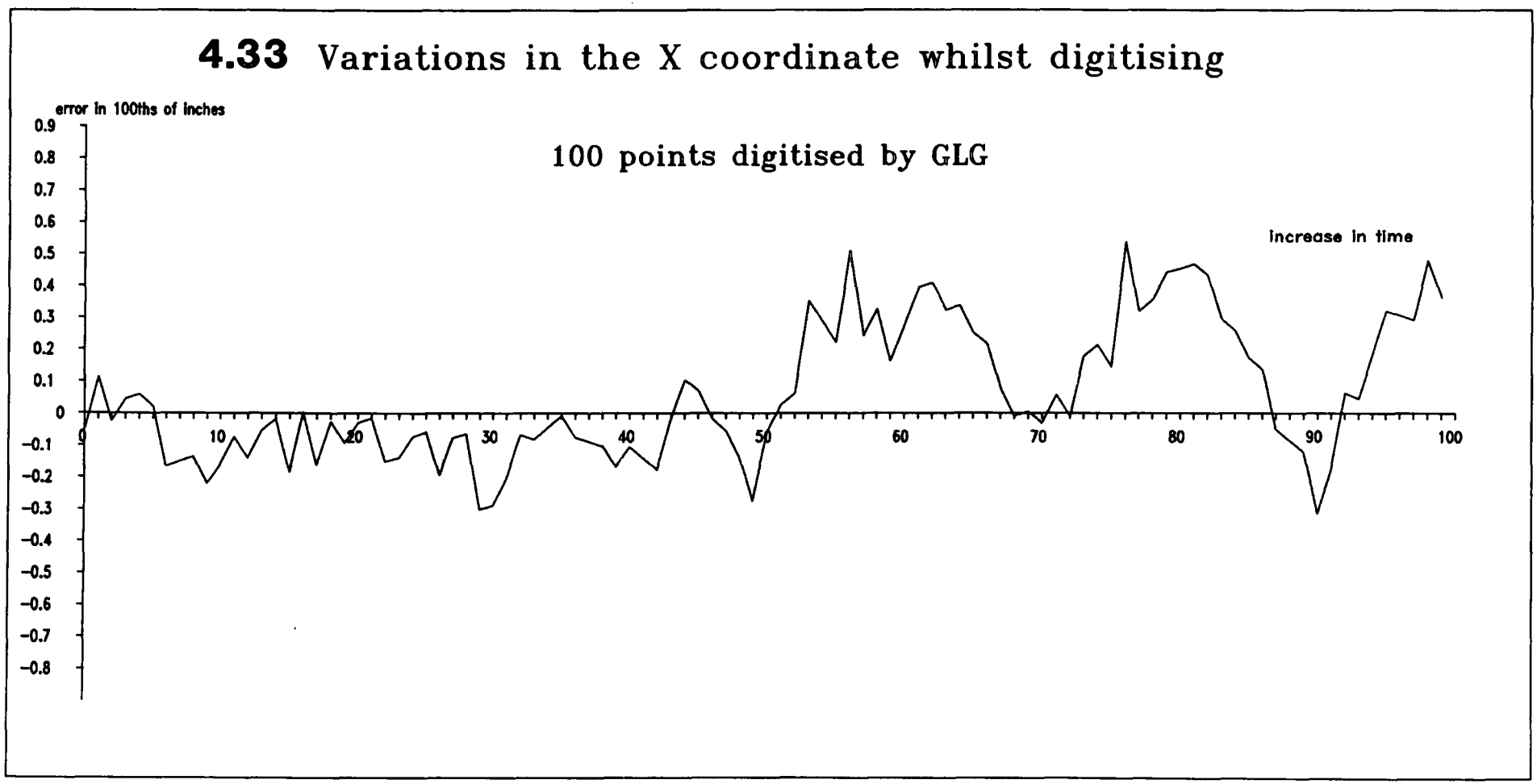



4.31 Variations in the Y-coordinate whilst digitising
error in 10 ths of inches
900 points digitised by KYR
increase in time
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7

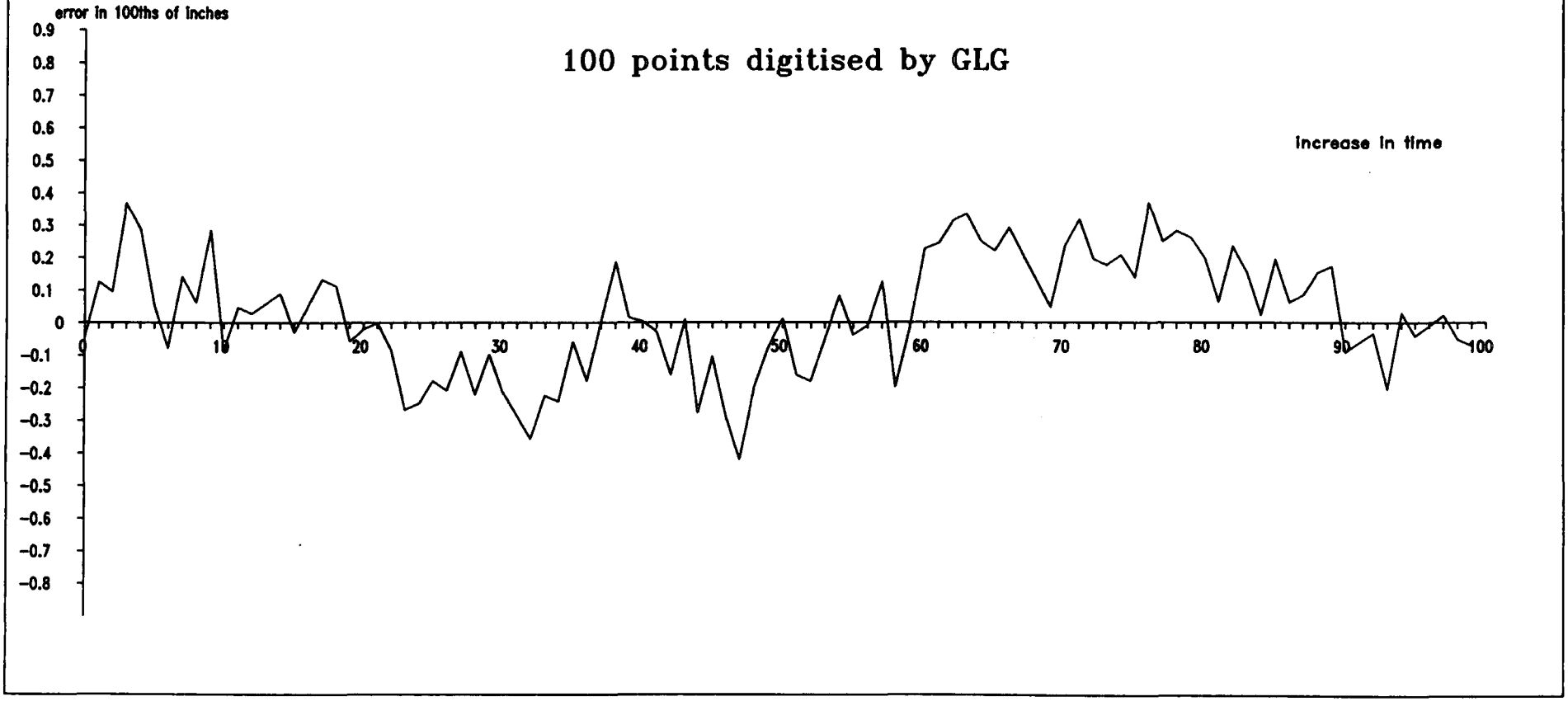
432 Variations in the Y coordinate whilst digitising



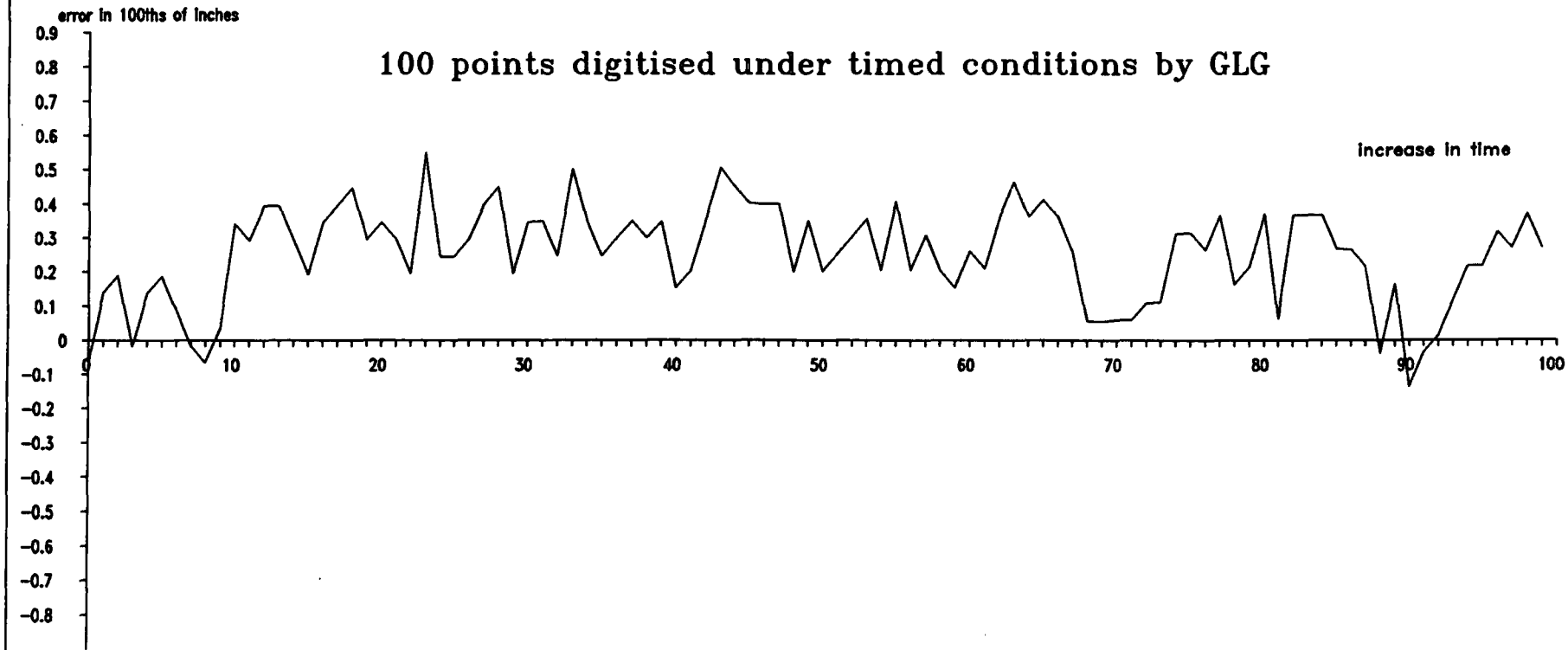
4.33 Variations in the X coordinate whilst digitising



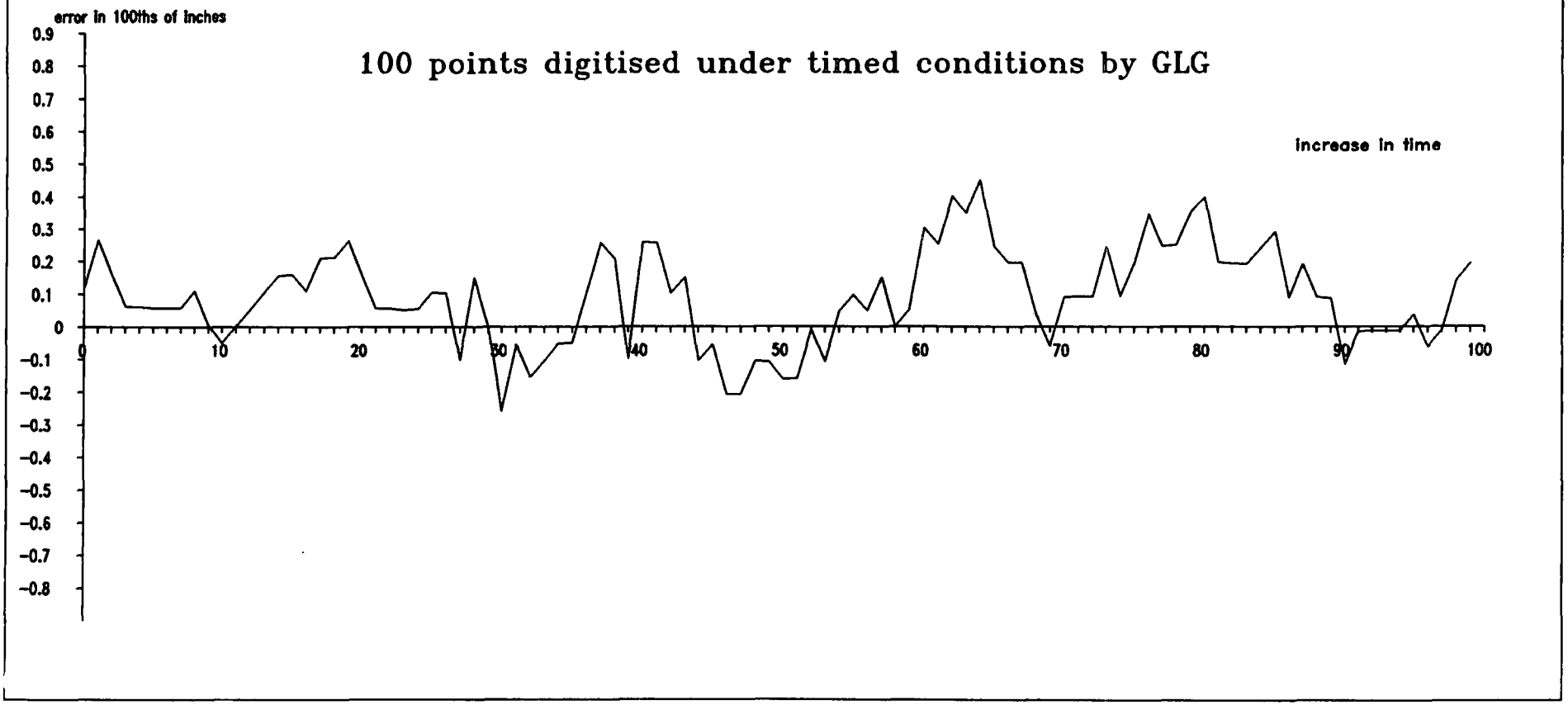
4.34 Variations in the Y coordinate whilst digitising

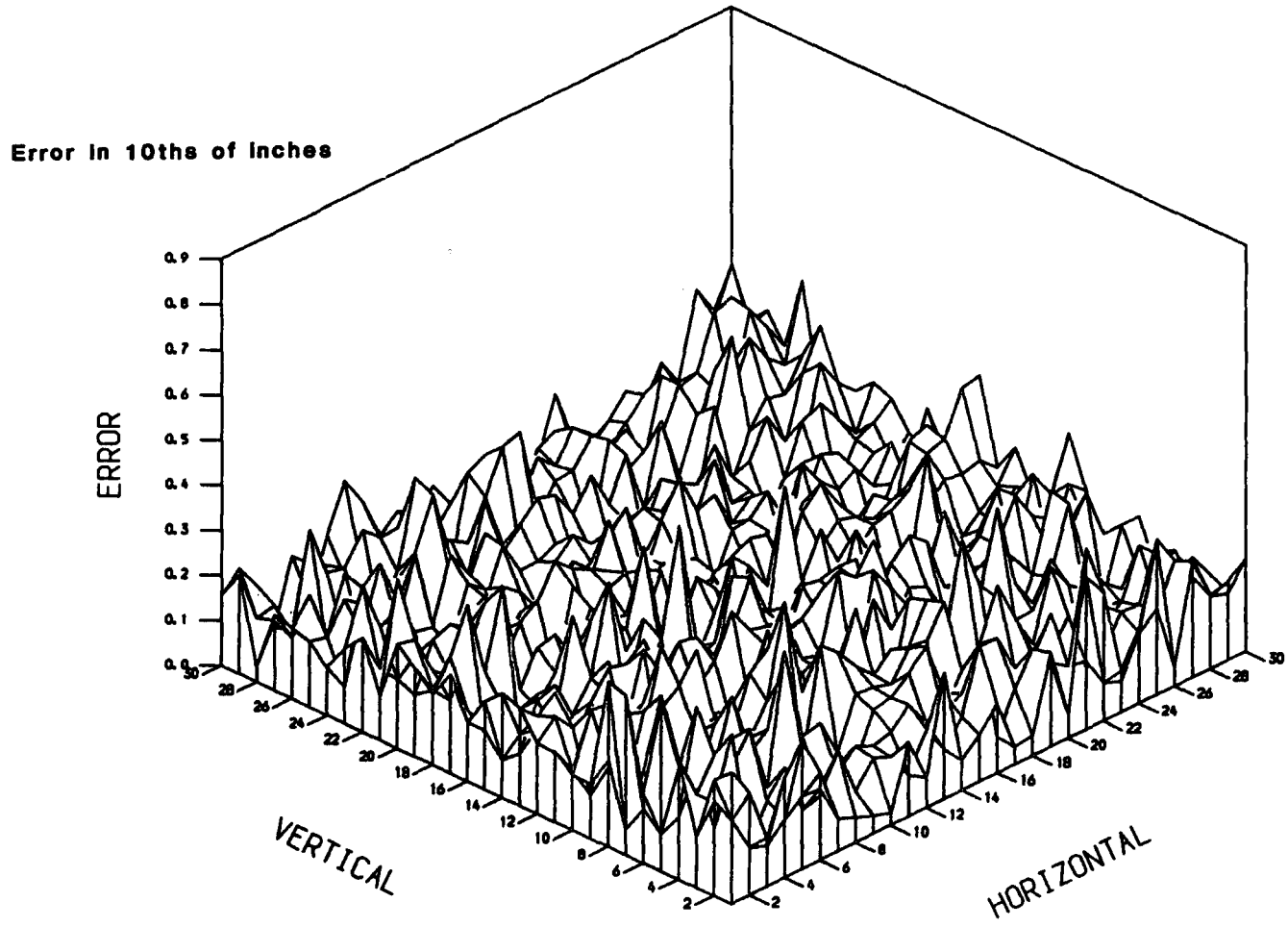


4.35 Variations in the X coordinate whilst digitising

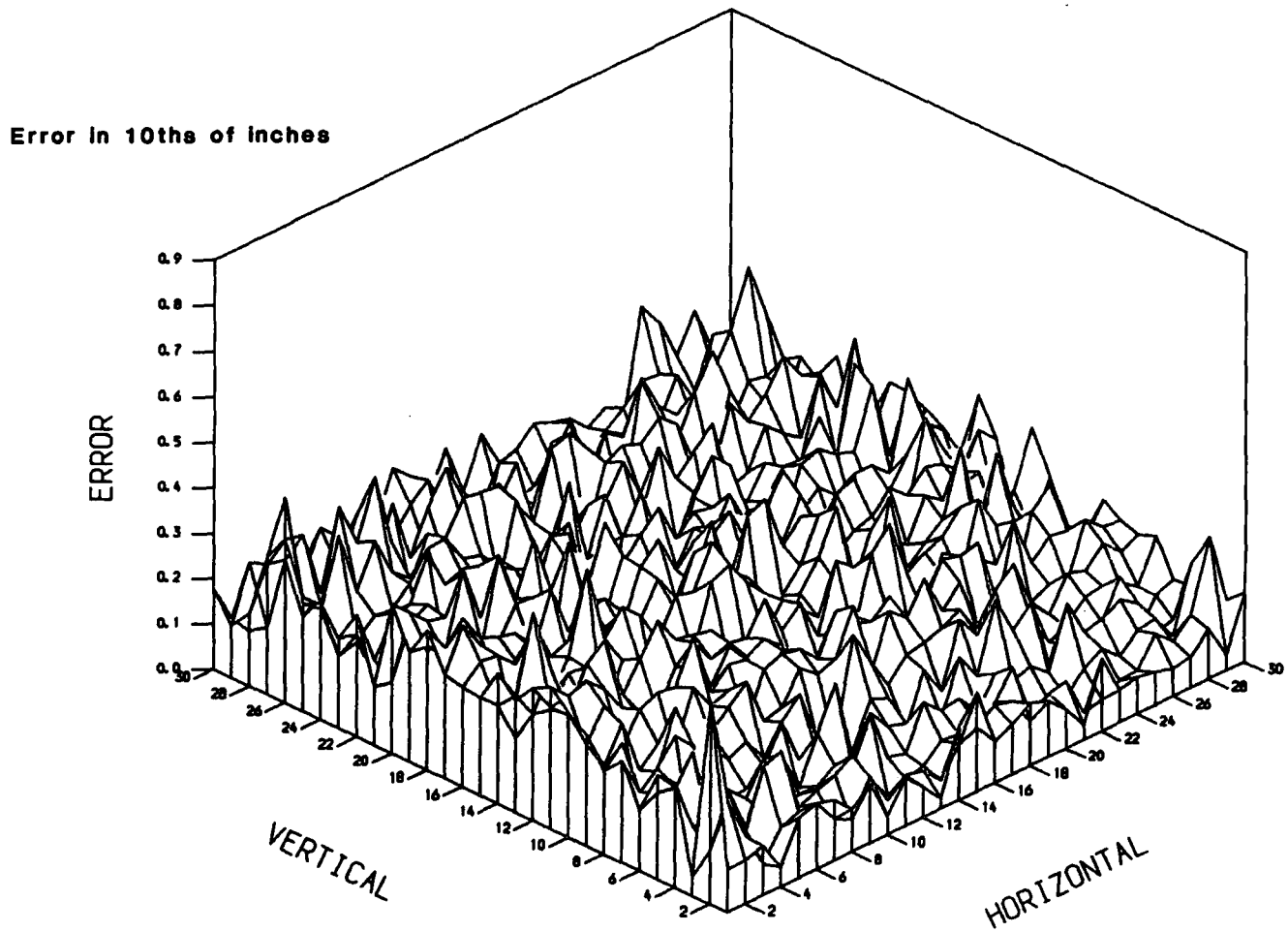


4.36 Variations in the Y coordinate whilst digitising

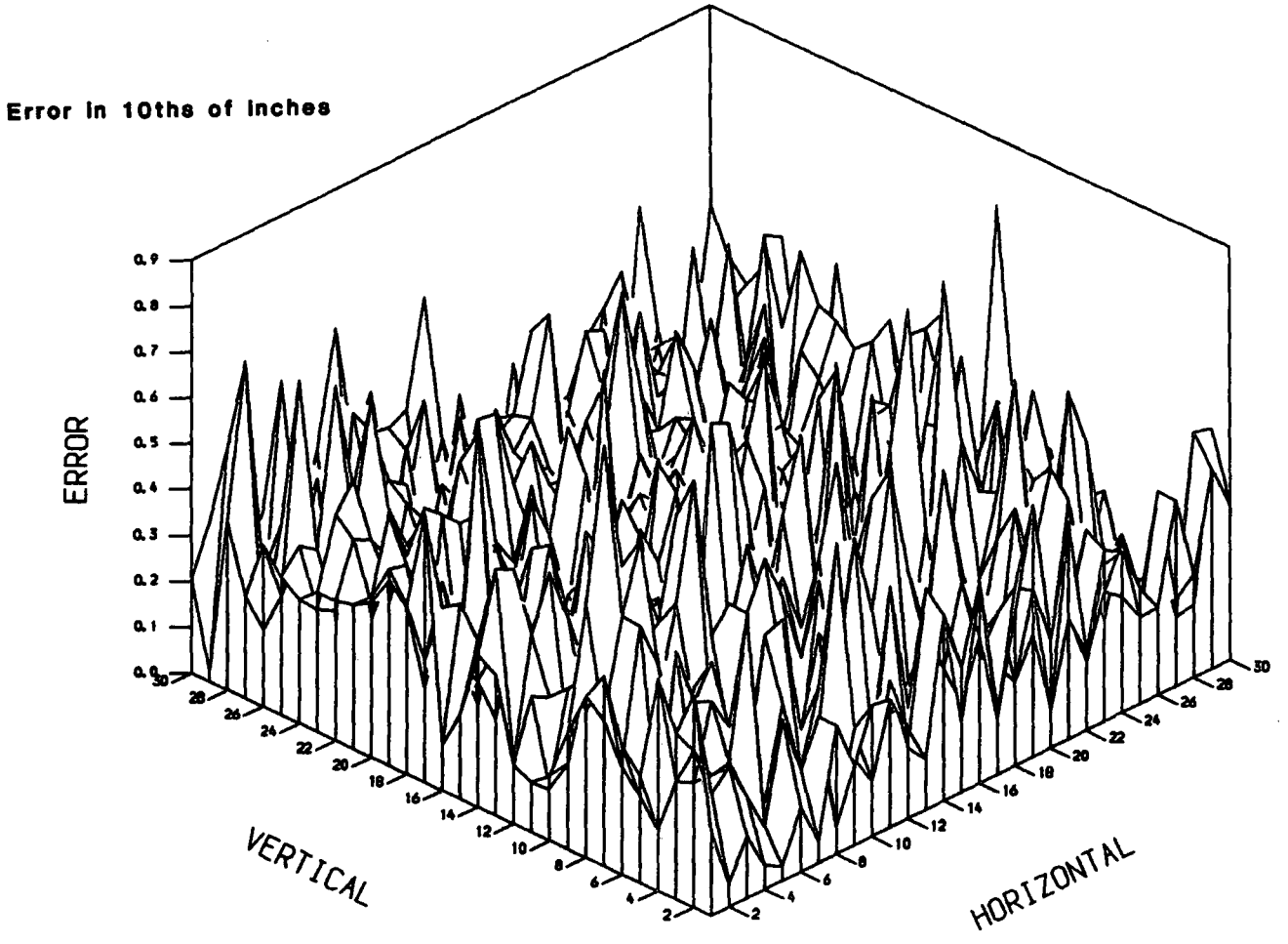




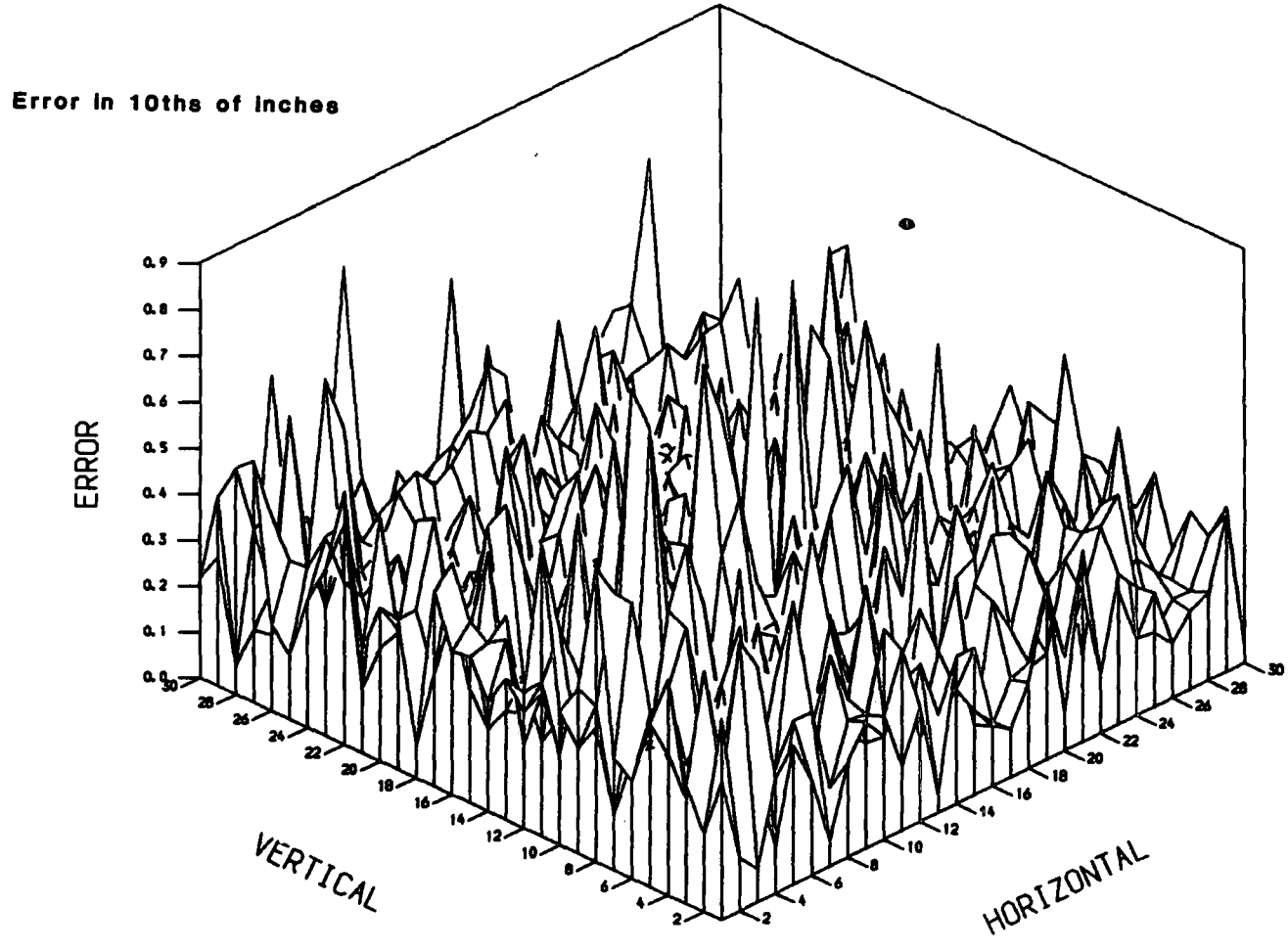
4.37 Digitising error with respect to the spatial extent of the digitsing table for operator NJP



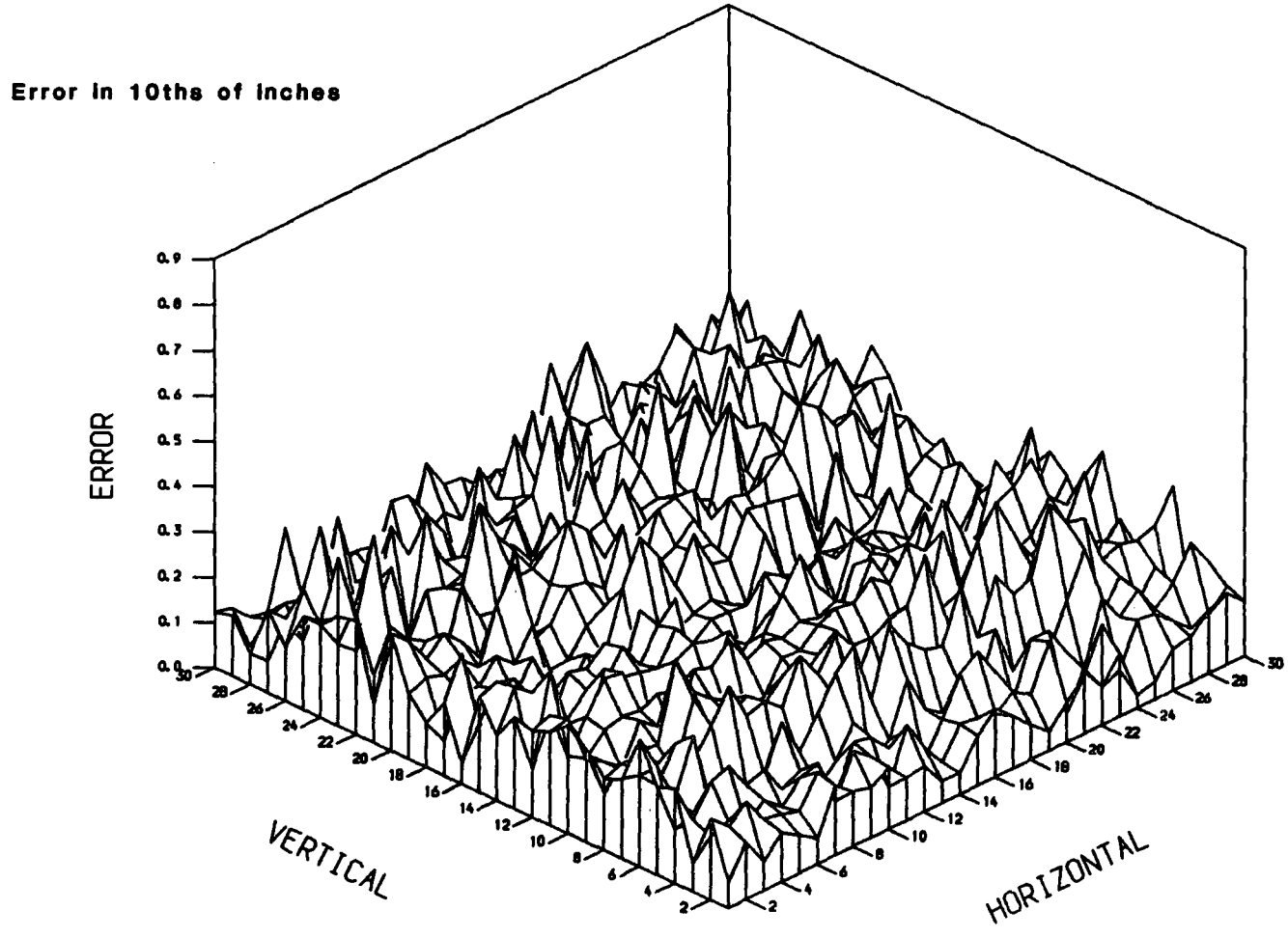
4.38 Digitising error with respect to the spatial extent of the digitsing table for operator CEH



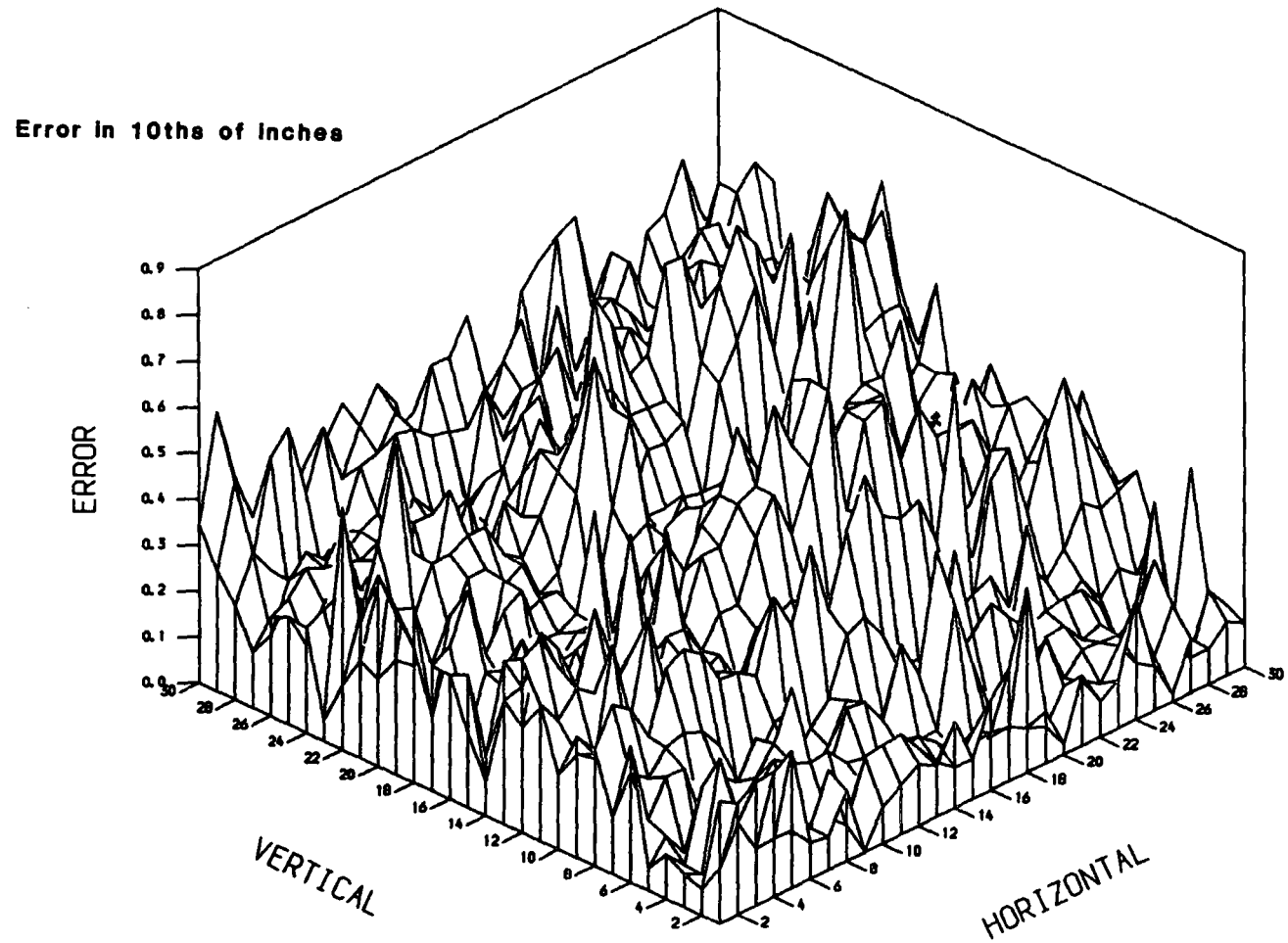
4.39 Digitising error with respect to the spatial extent of the digitsing table for operator KYR



4.40 Digitising error with respect to the spatial extent of the digitsing table for operator MJB



4.41 Digitising error with respect to the spatial extent of the digitsing table for operator GLG



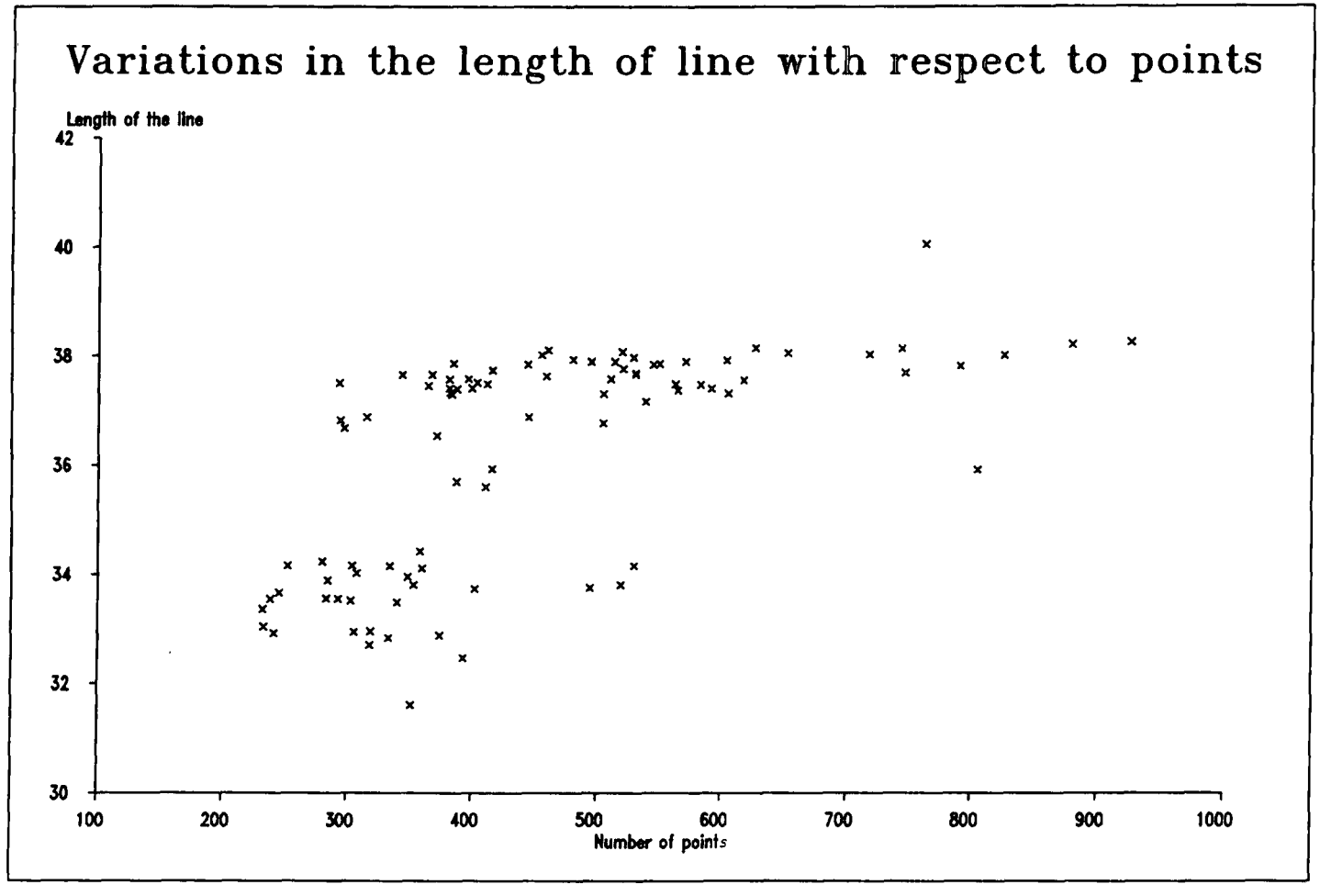
4.42 Digitising error with respect to the spatial extent of the digitsing table for operator ALS

4.43

A frequency histogram of the length of lines when digitised from different scales using different operators.

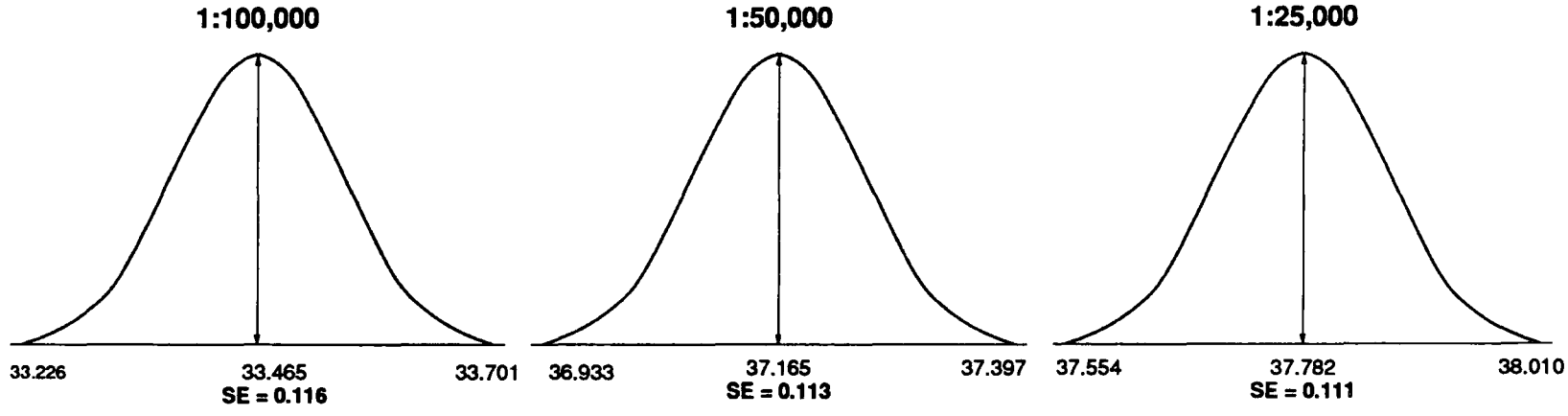
```
COUNT    MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY   .50 OCCURRENCES

    0       30.8
    0       31.3
    1       31.8  **
    1       32.3  **
    7       32.8  **************
    6       33.3  ************
    8       33.8  ****************
    7       34.3  **************
    0       34.8
    1       35.3  **
    3       35.8  ******
    1       36.3  **
    5       36.8  **********
   20       37.3  ****************************************
   24       37.8  ************************************************
    5       38.3  **********
    0       38.8
    0       39.3
    1       39.8  **
    0       40.3
    0       40.8
                  I____|____I____|____I____|____I____|____I____|____I____|____I
                  0         5         10        15        20        25
                            HISTOGRAM FREQUENCY
```

Figure 4.44

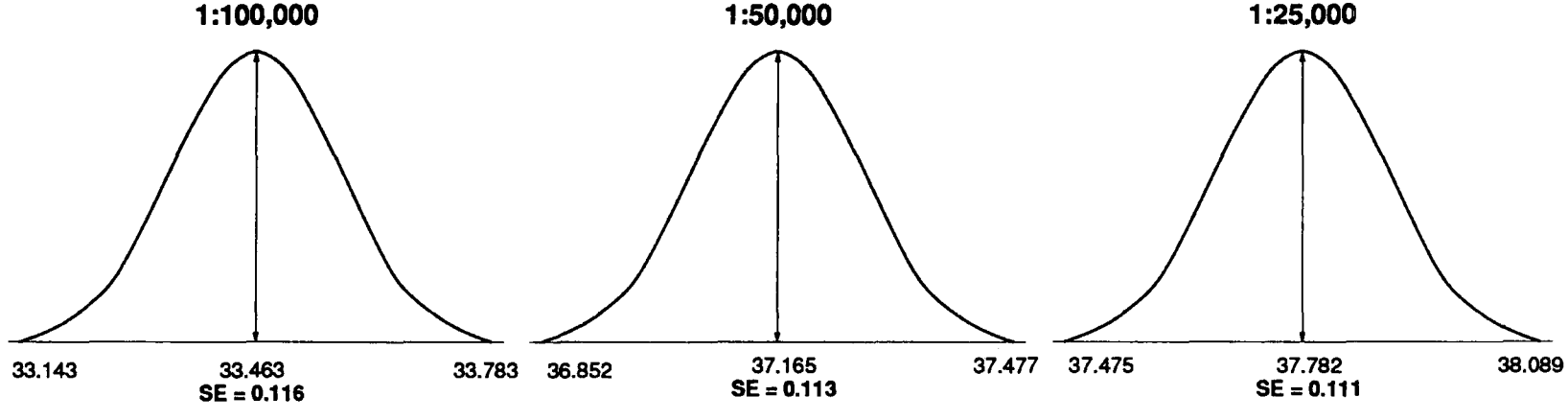

Variations in the length of line with respect to points
Length of the line
42
40
38
36
34
32
30
100
200
300
400
500
600
700
800
900
1000
Number of points

4.45 CONFIDENCE LIMITS FOR THE SIMILARITY OF DIGITISED LINES

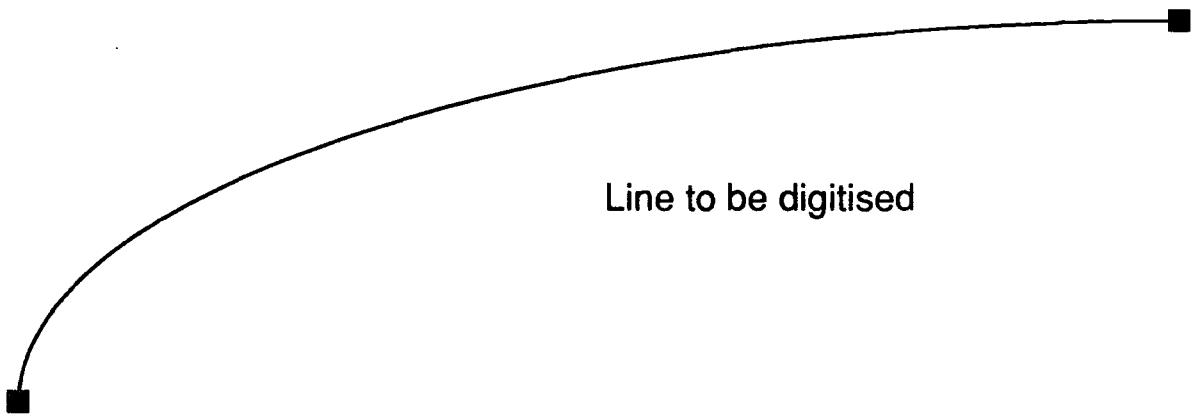
AT 95% SIGNIFICANCE



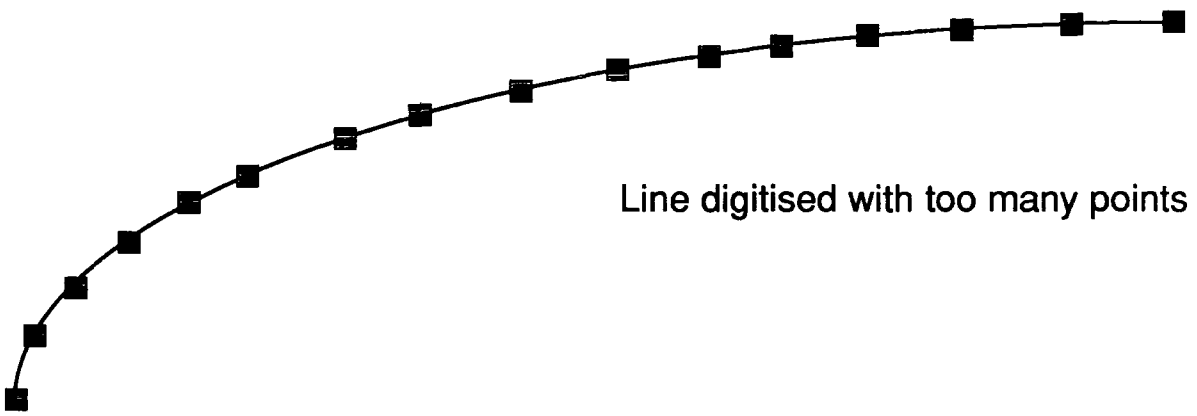
AT 99% SIGNIFICANCE



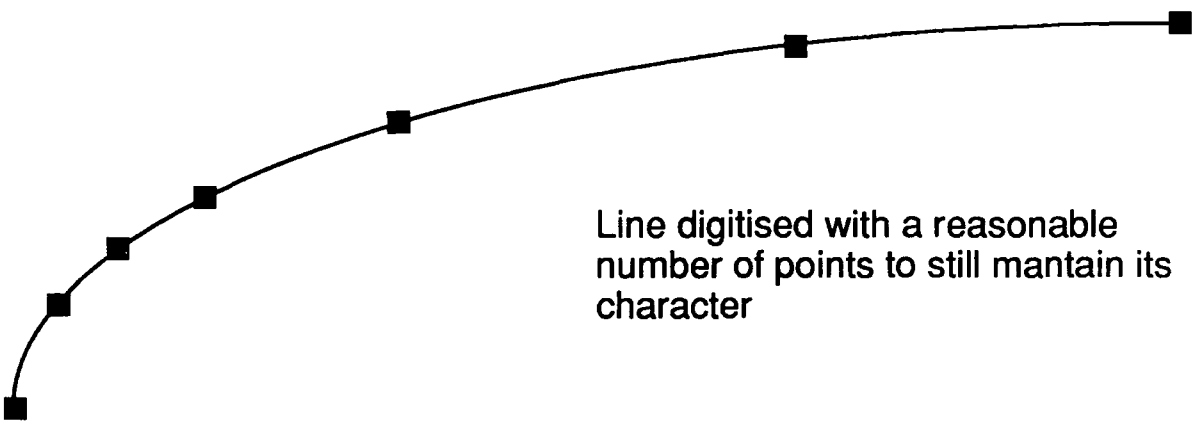
4.46 Using too many points to efficiently characterise a line



Line to be digitised



Line digitised with too many points

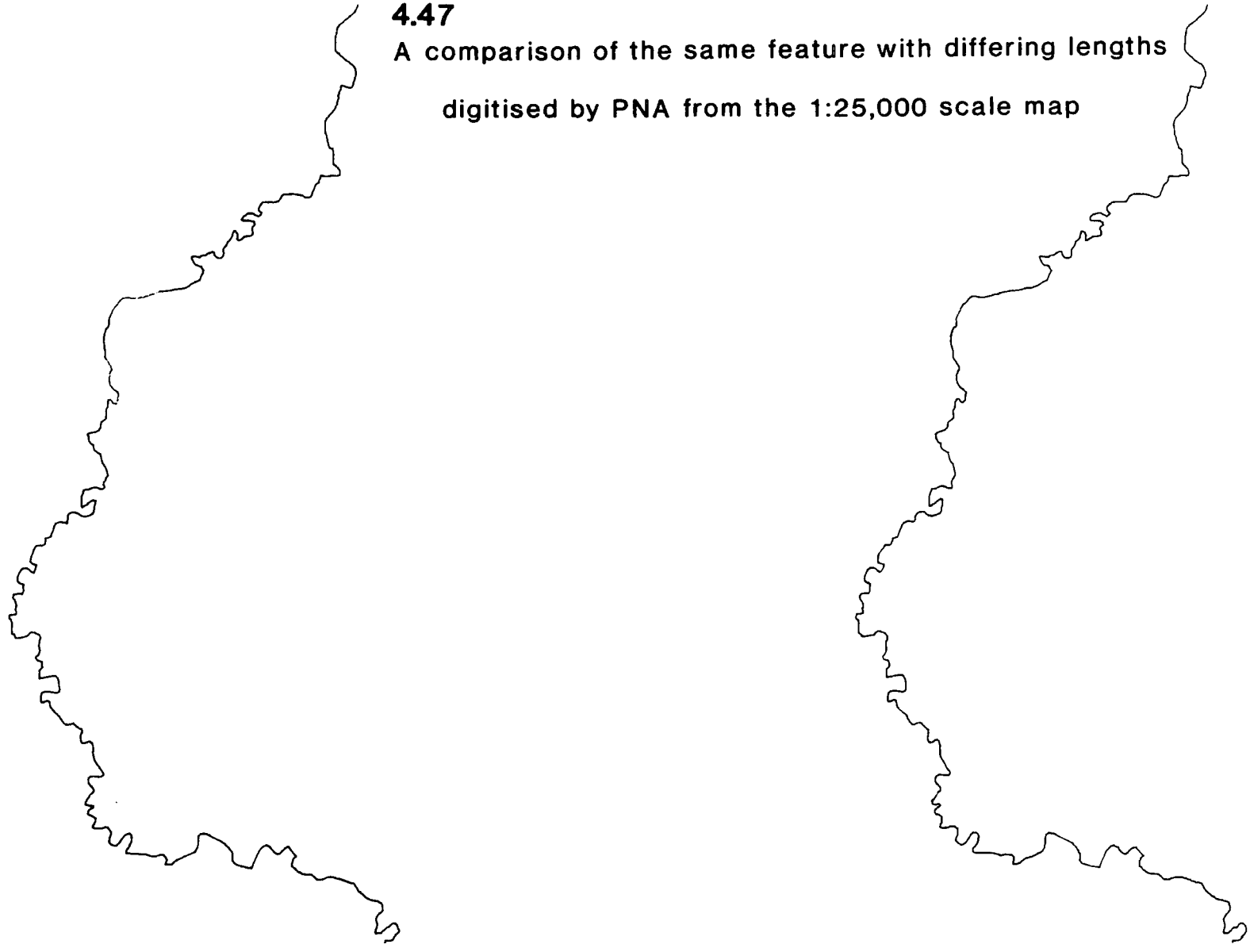


Line digitised with a reasonable number of points to still mantain its character

4.47

A comparison of the same feature with differing lengths

digitised by PNA from the 1:25,000 scale map

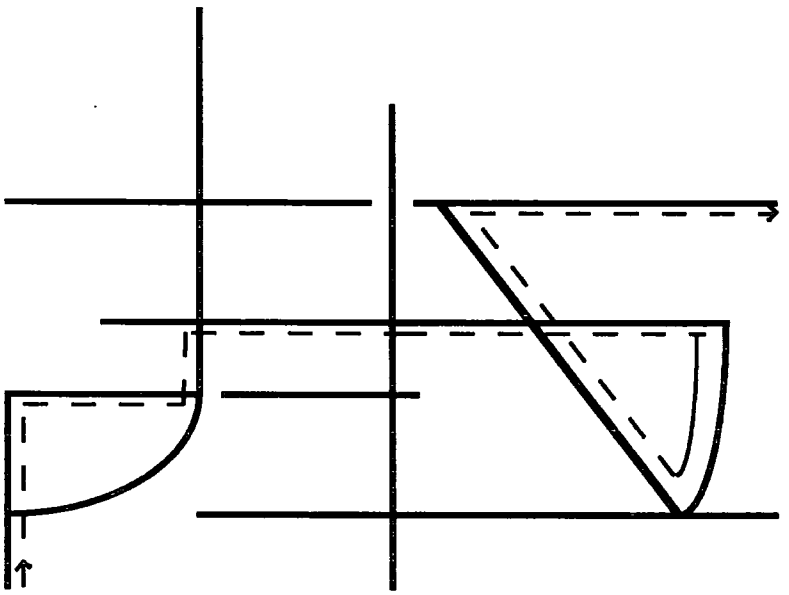


Total line length: 40.0032km

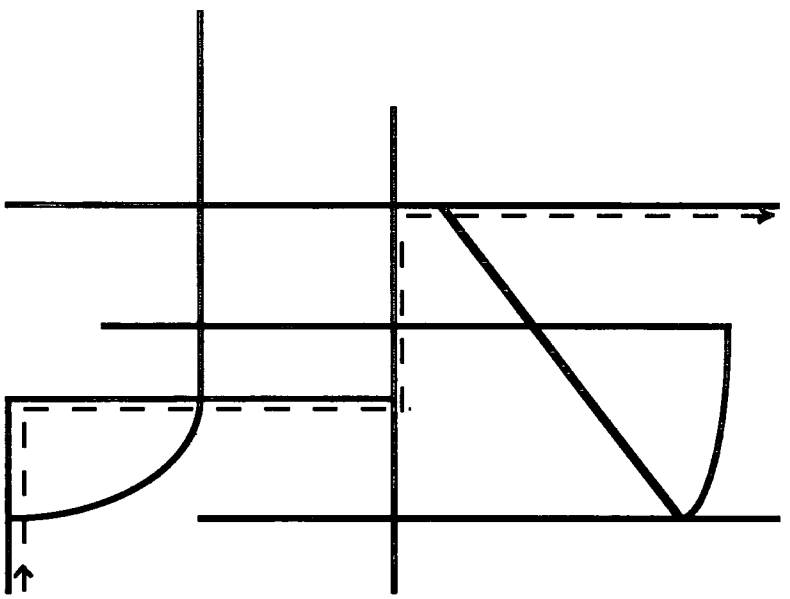
Total line length: 35.8605km

Figures for Chapter 5

5.1. Network structured data illustrating the need for contiguity so that flows can be modelled

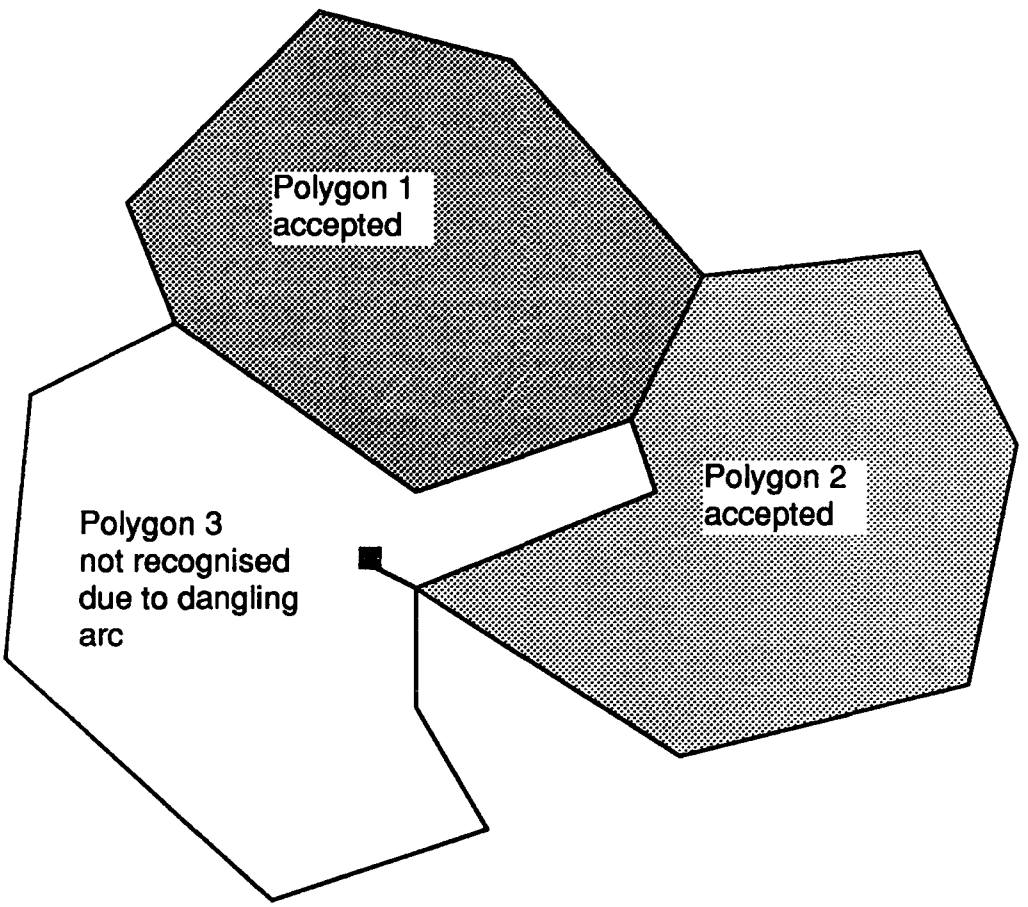


Network unable to function using a shortest route algorithm due to the discontinuous nature of the digitised lines

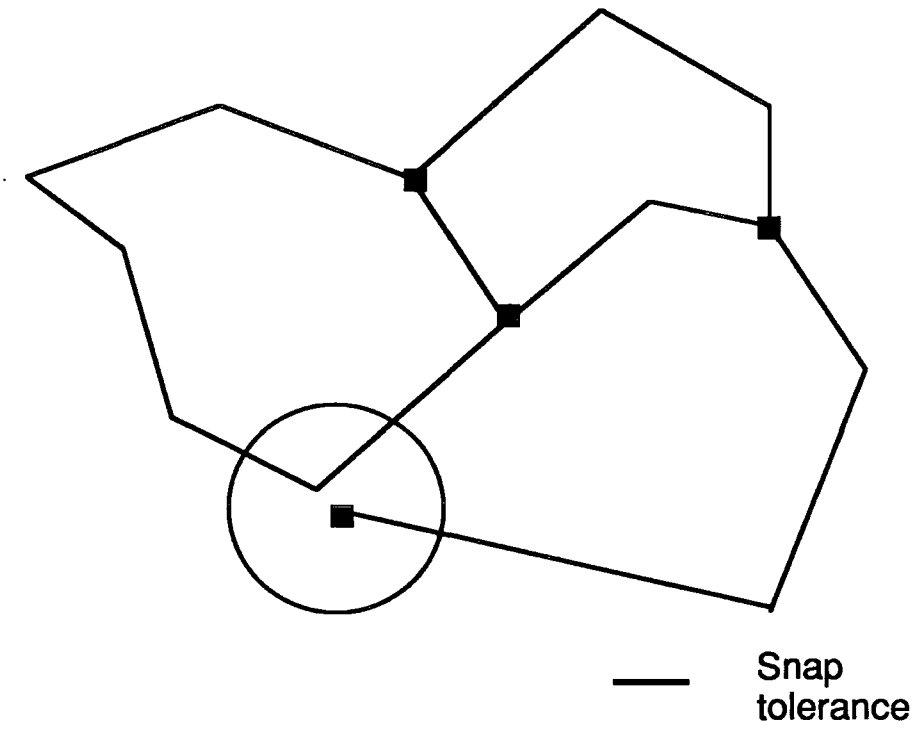


Network selects shortest route as all lines are contiguous

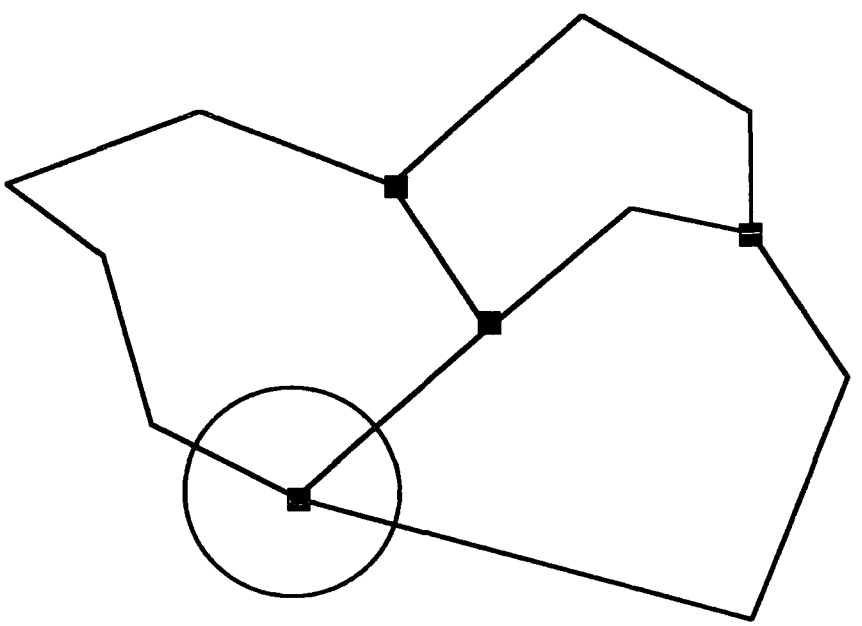
5.2 Data model unable to create full polygon structure due to the presence of a dangling arc



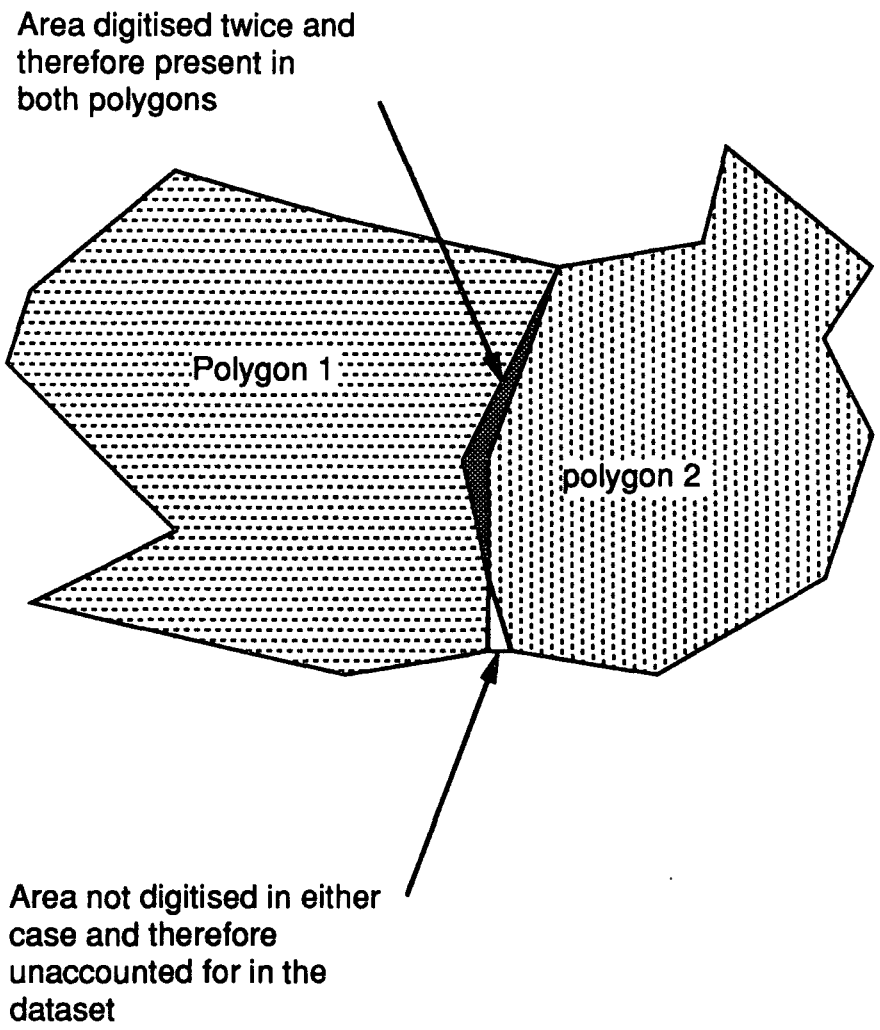
5.3 Snap tolerance used in digitising



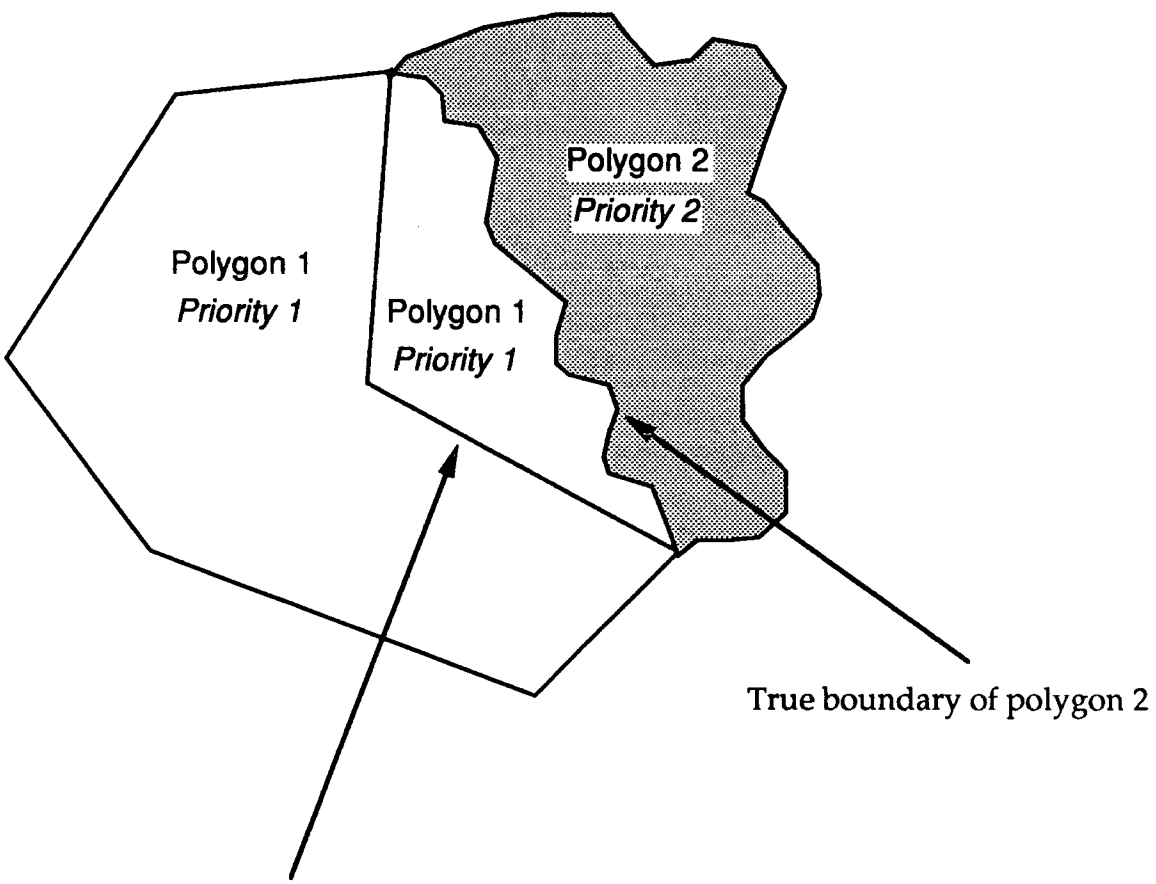
Snap tolerance allows the line to be 'snapped' onto the nearest available node, thus the operator merely has to end the line NEAR the node at which it should finish



5.4. Entity by entity digitising and possible problems

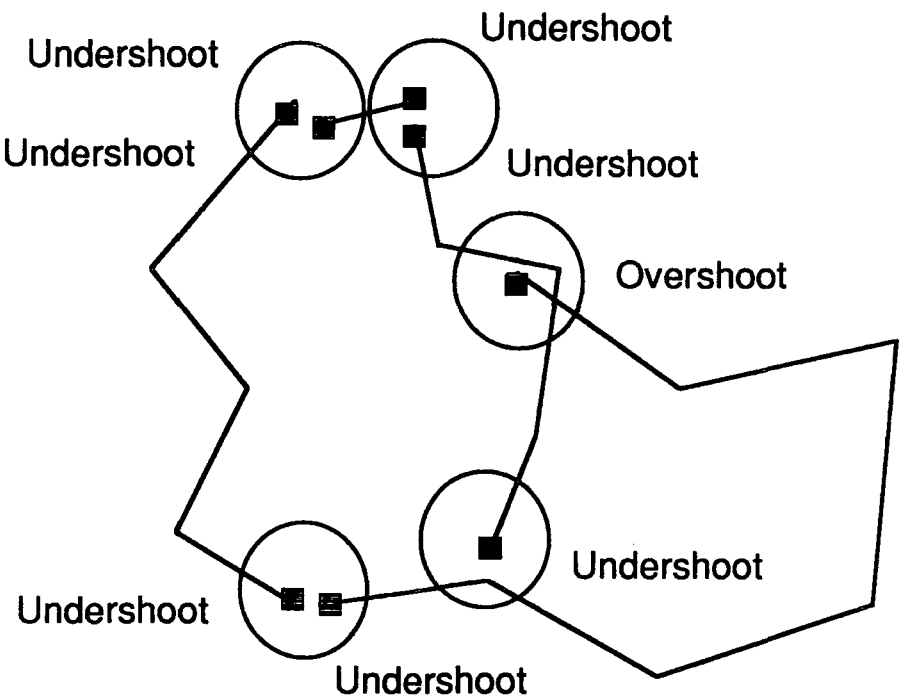


5.5 Priority numbering of polygons as an alternative to having to digitise complex lines twice



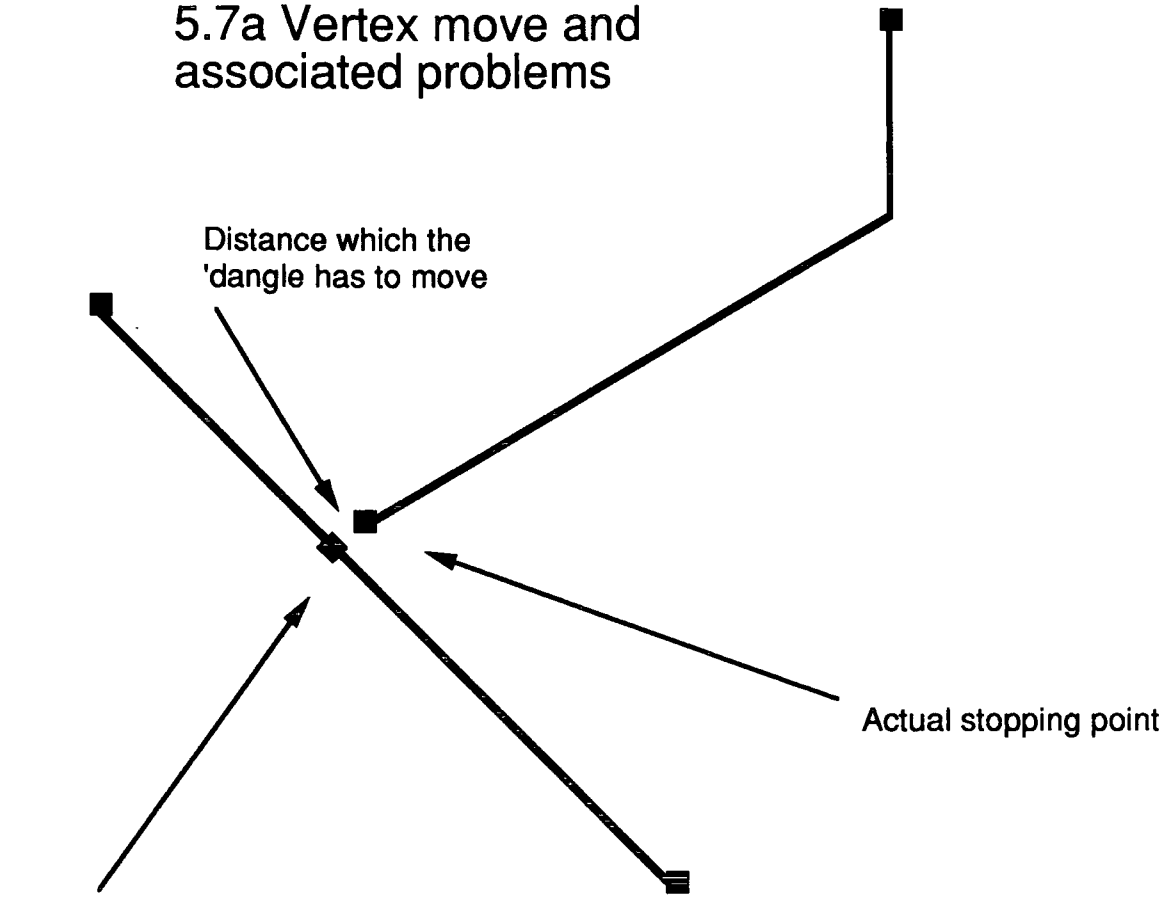
Falsely extended boundary of polygon 2, which will be removed by the internal processing, due to the fact that polygon 2 has a lower priory number to polygon 1. In this case, as in all others, if there is a conflict, then the boundary of polygon 1 takes priority

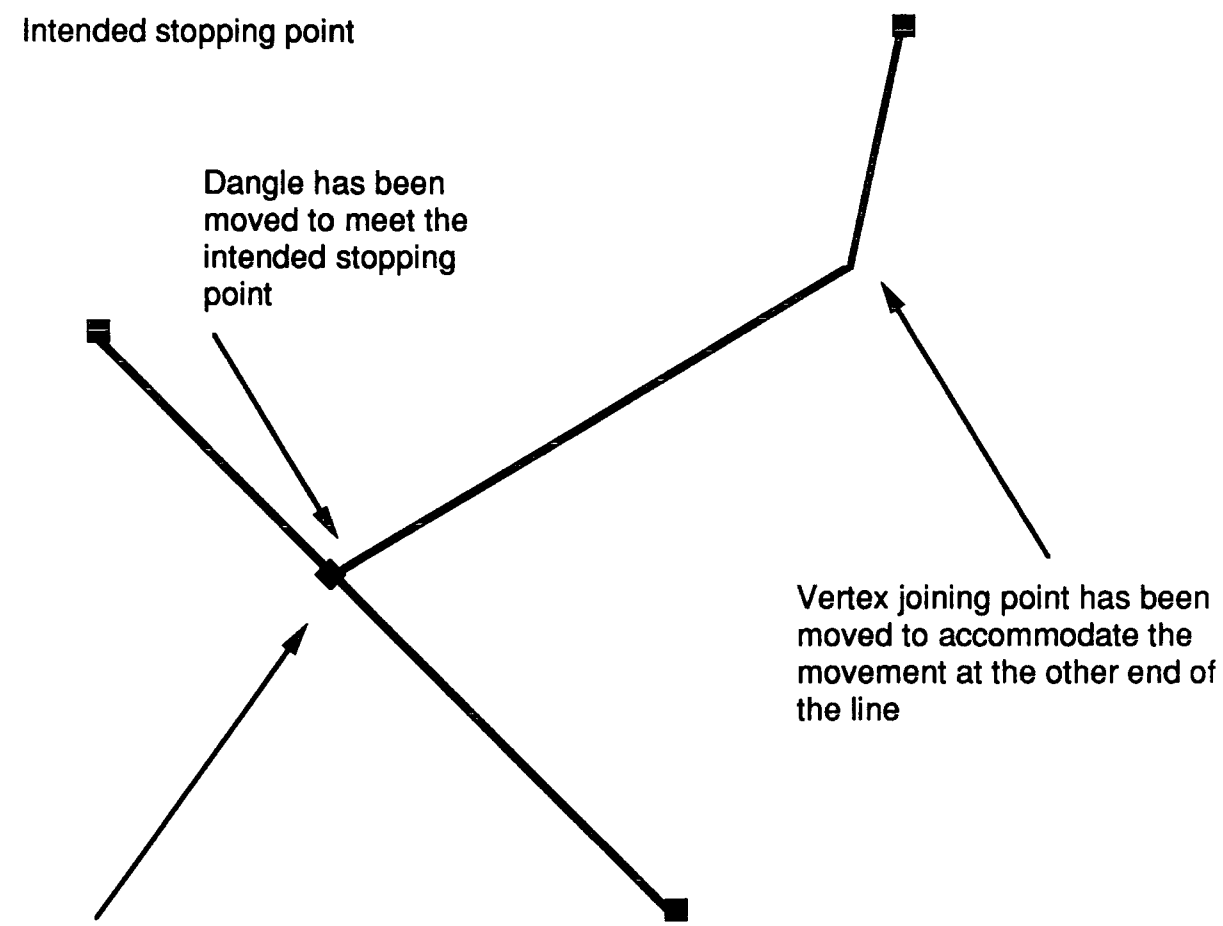
5.6 Undershoots and overshoots



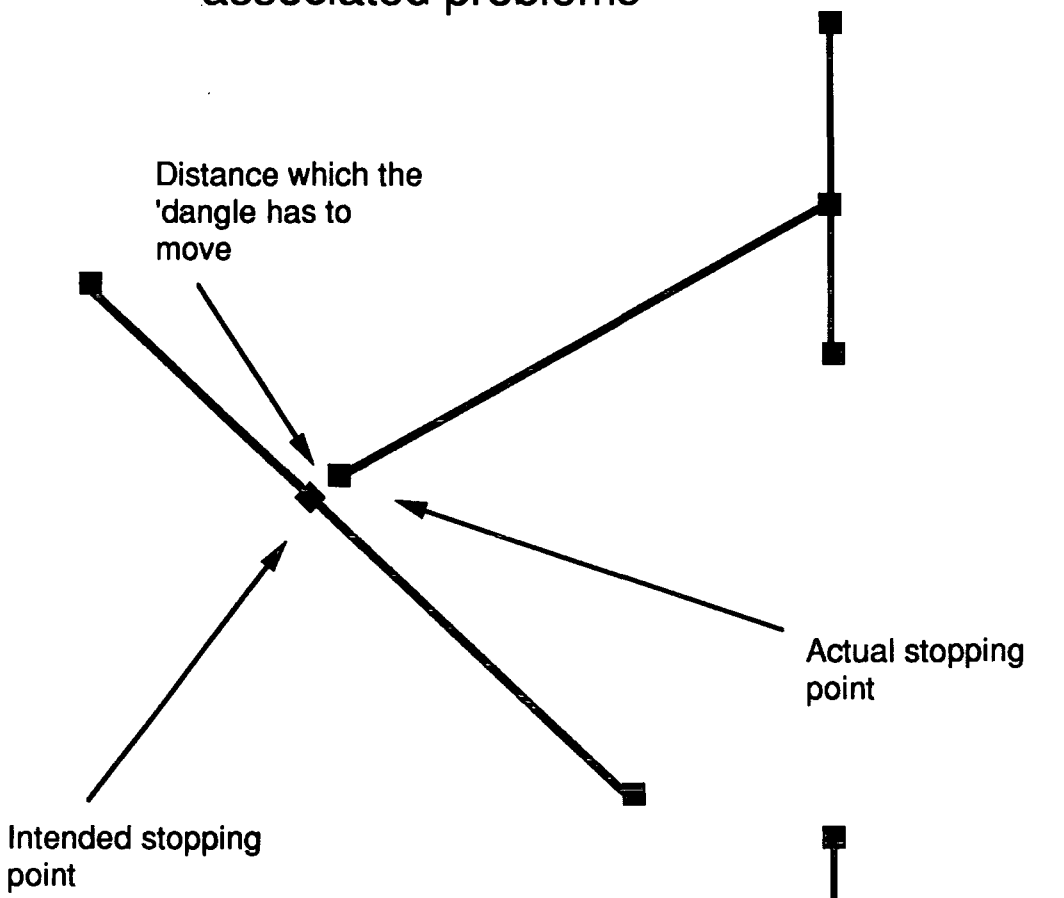
Undershoots and overshoots in the digitising process are both recognised as dangling arcs by the data model, as in both cases the lines onto which these lines are attached 'dangle' in space

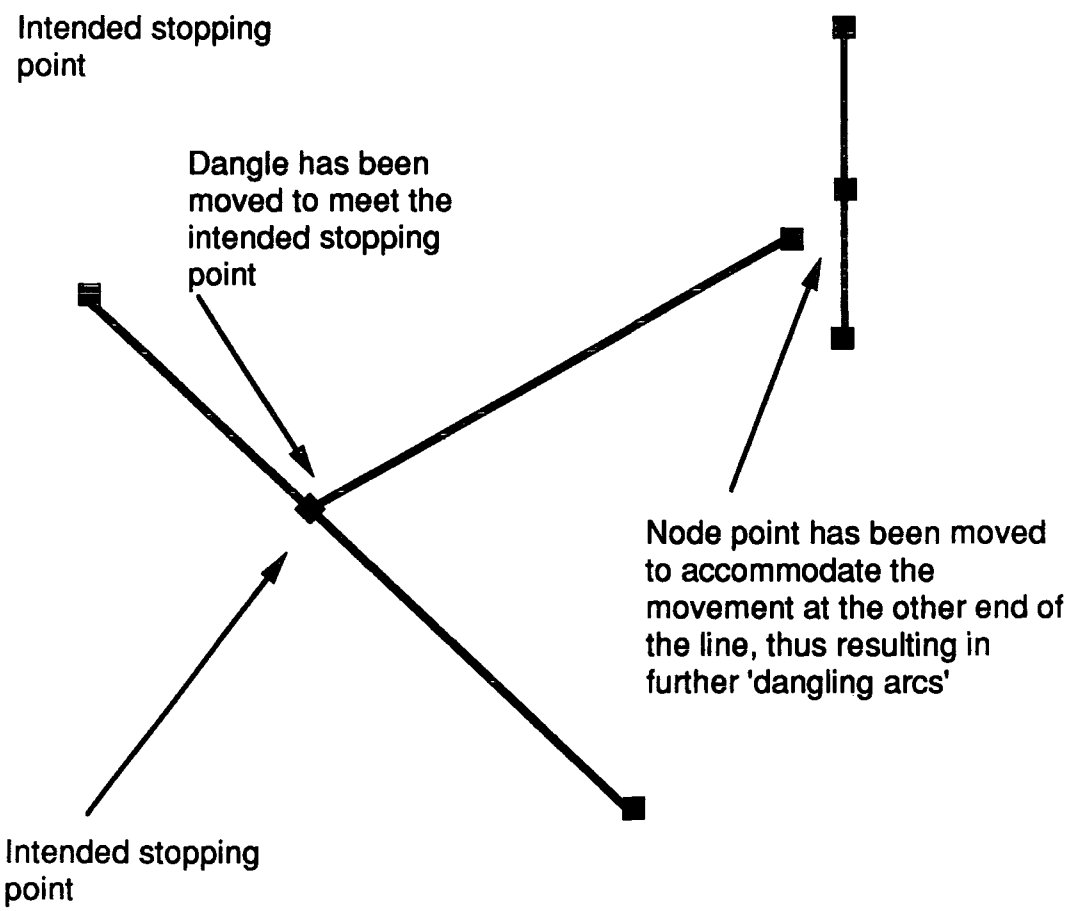
5.7a Vertex move and associated problems



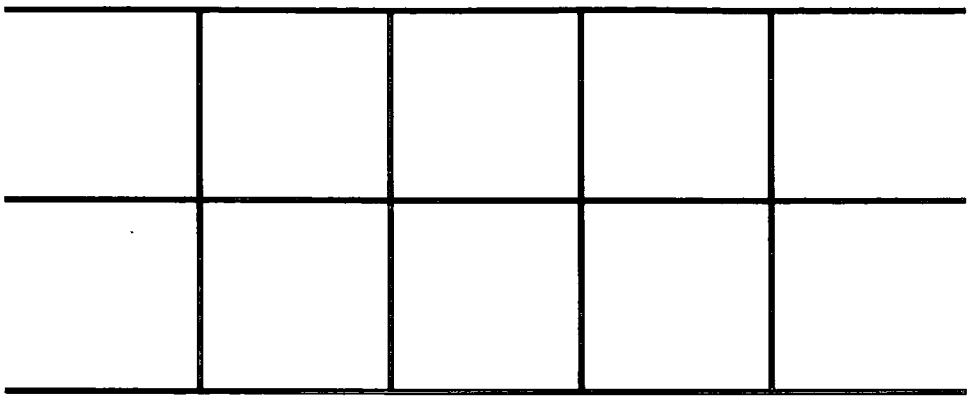


5.7b Arc move and associated problems

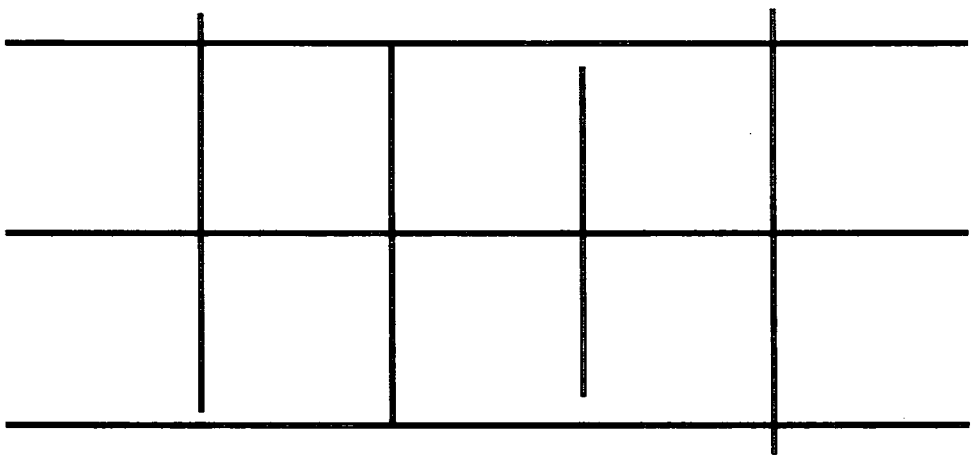




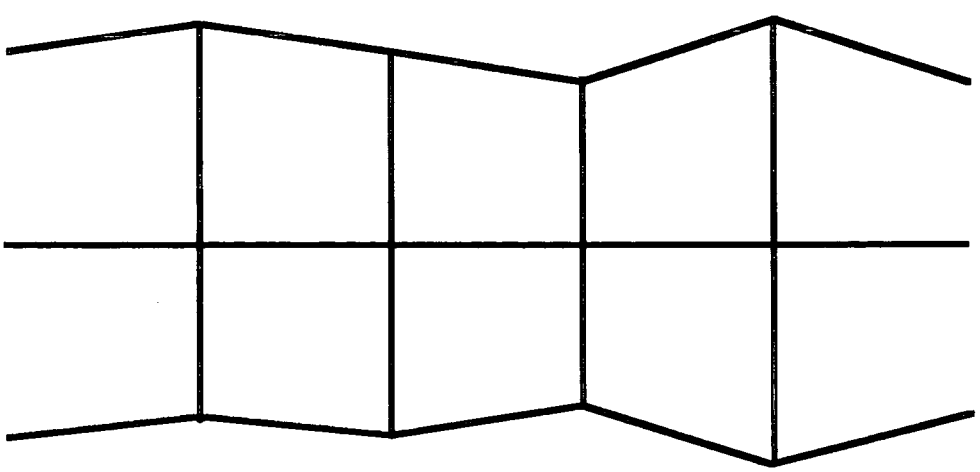
5.8 Problems of vertex movement using straight line coverages



The original features as they exist on the ground

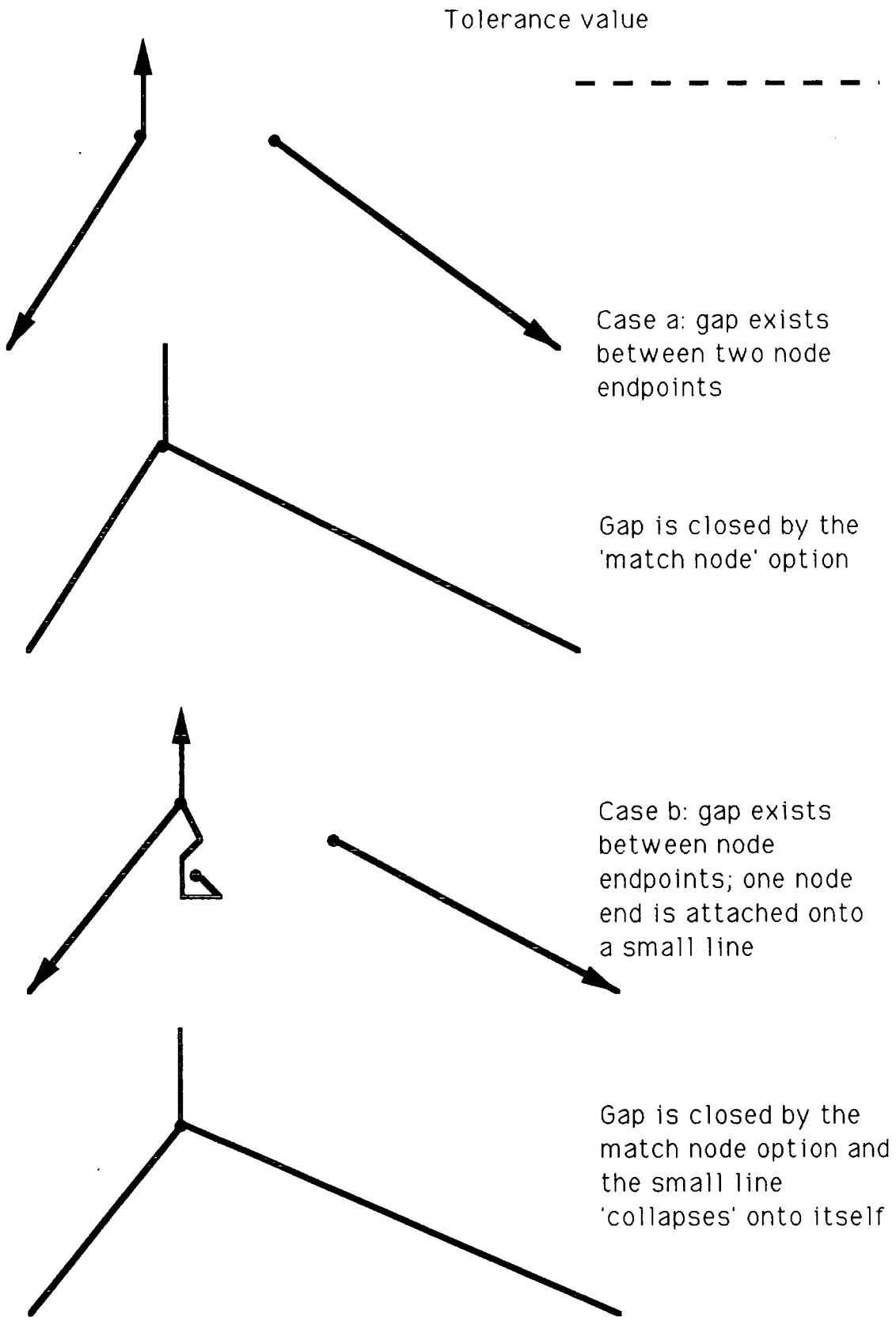


Lines have been digitised with a series of overshoots and undershoots



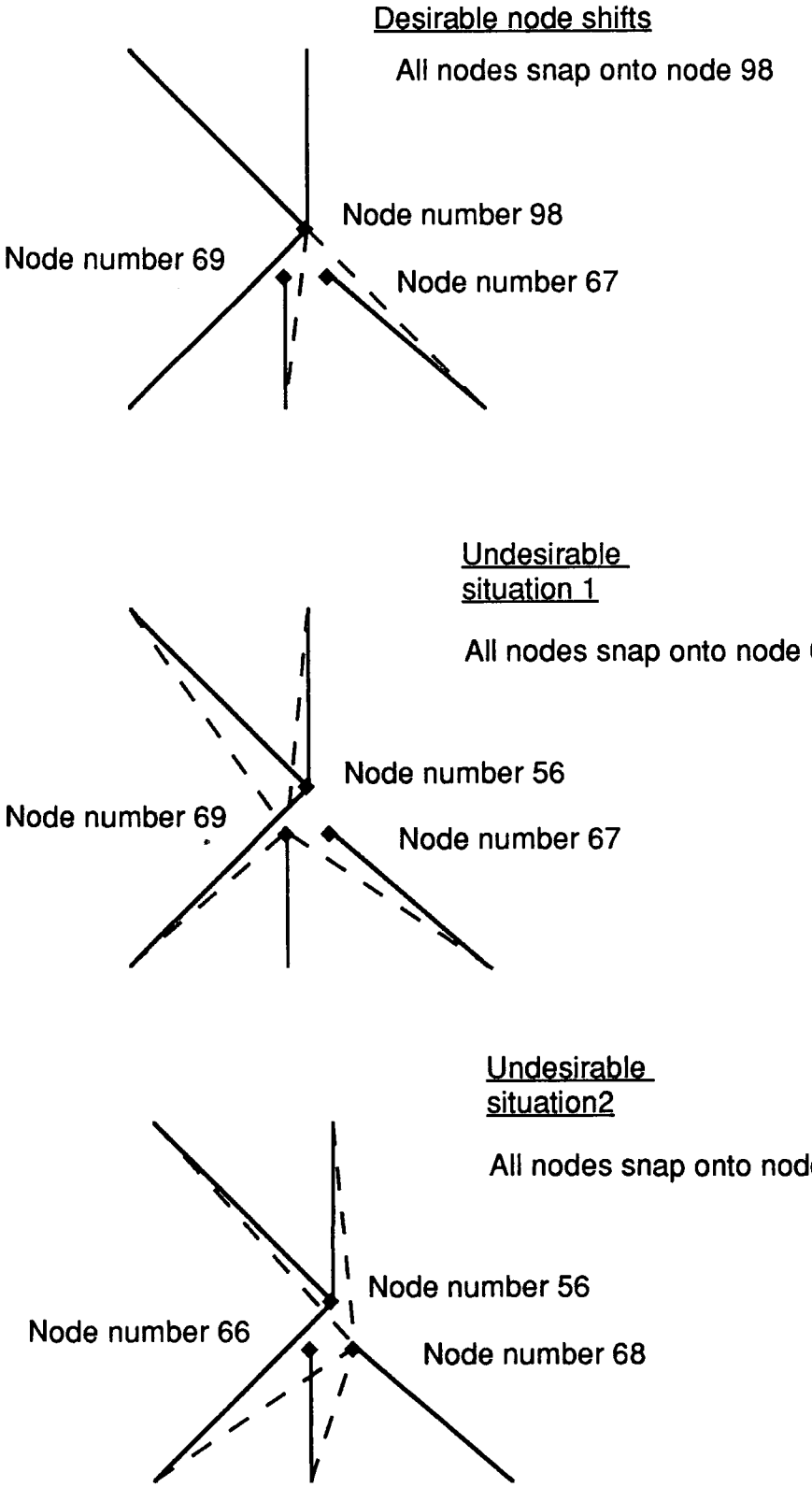
Worst possible scenario, where accurate vertices are moved to meet dangling arcs

5.9 Problems of node collapse

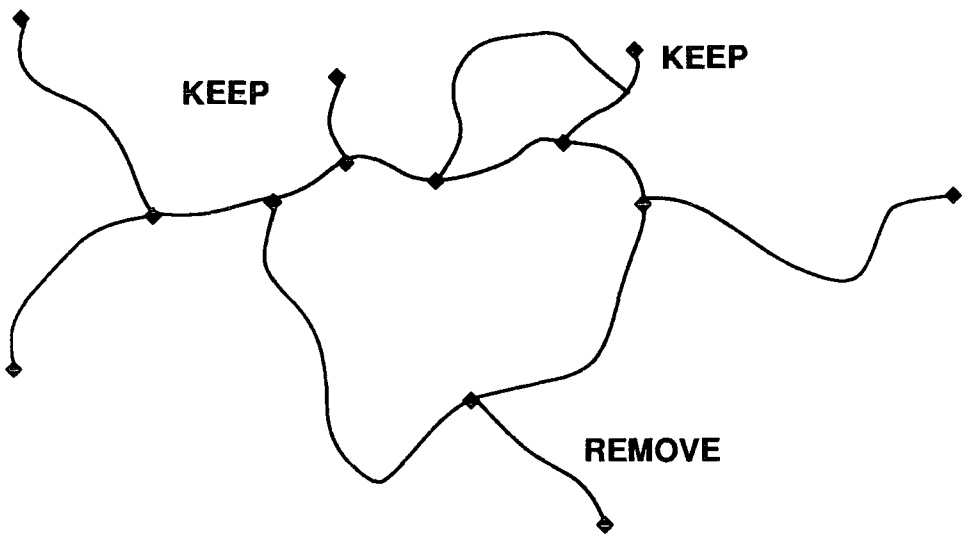


5.10 The unpredictability of node matches as a consequence of random node numbering.

The error tolerance is applied to each node in numerical sequence, any neighbouring nodes which fall within the tolerance are snapped onto this node.

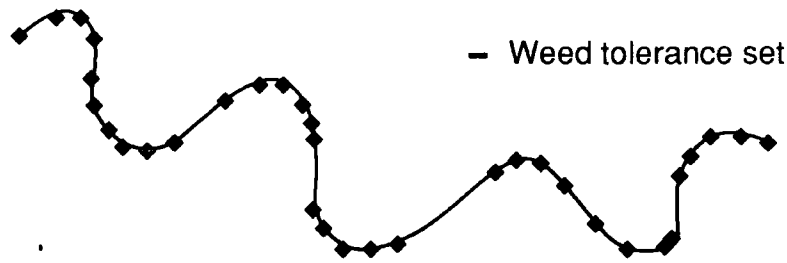


5.11 The removal of intentional 'dangling' lines in an attempt to remove node errors

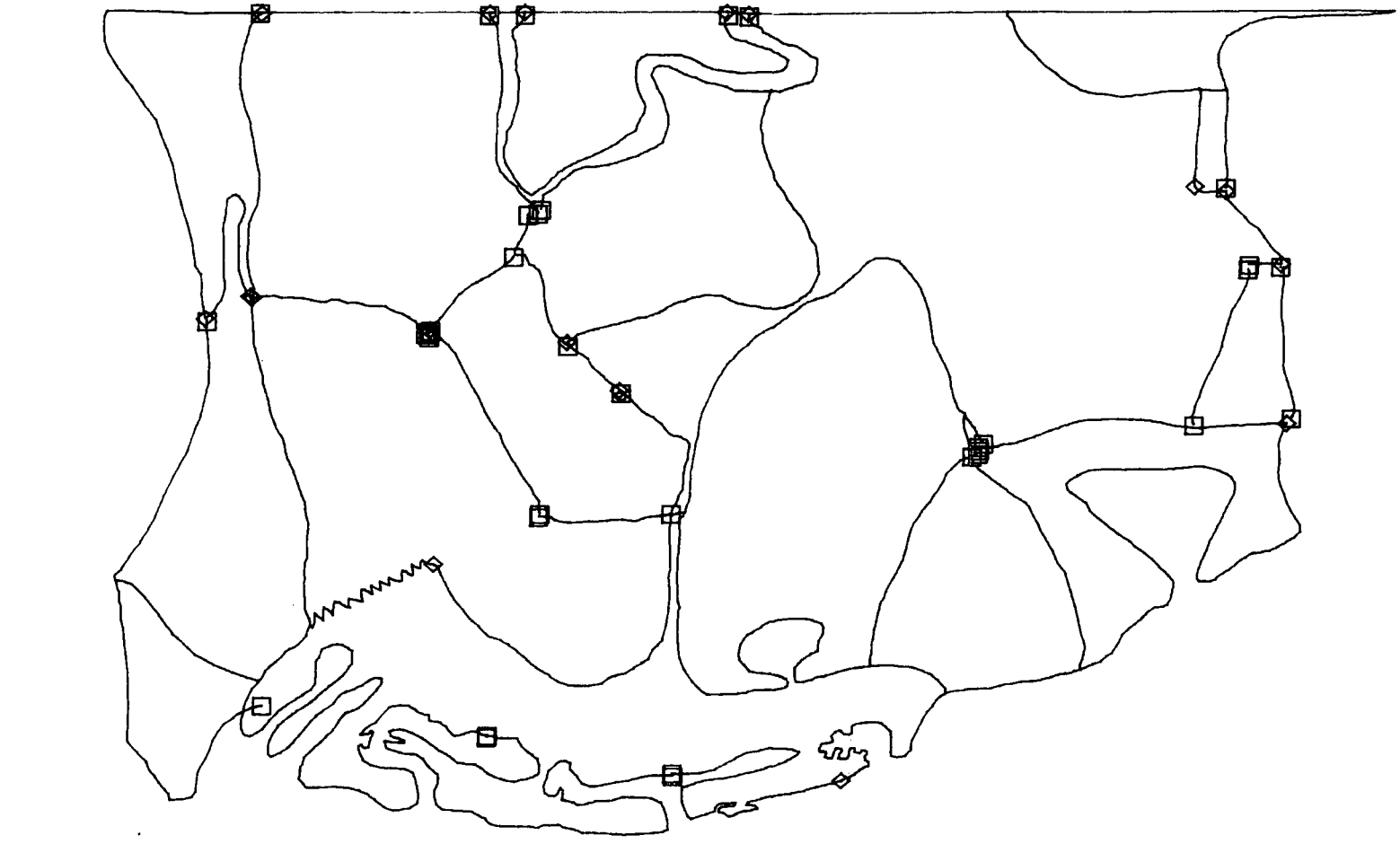


Tolerance set in order to remove large unwanted line

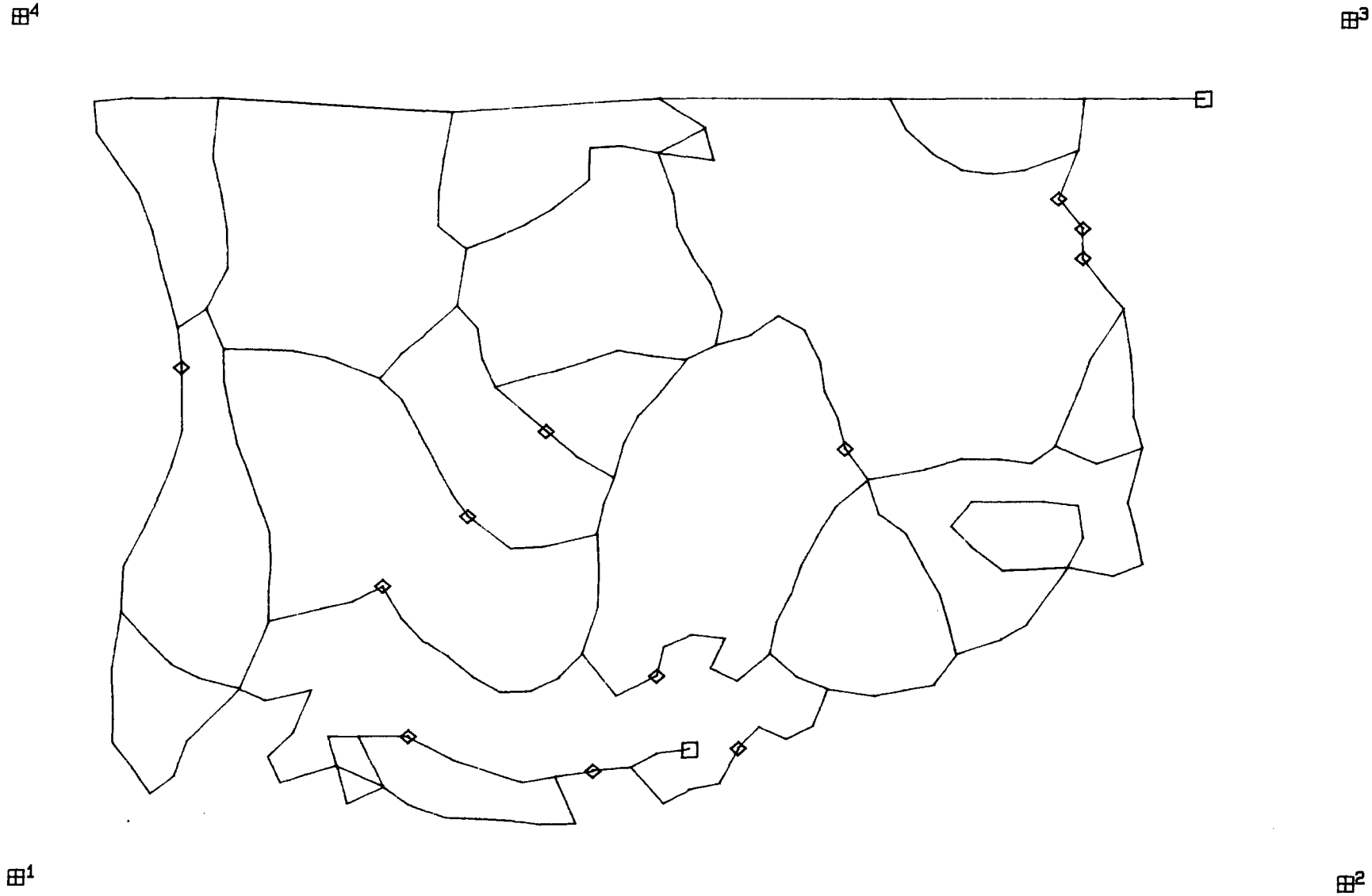
5.12 The 'weeding' procedure



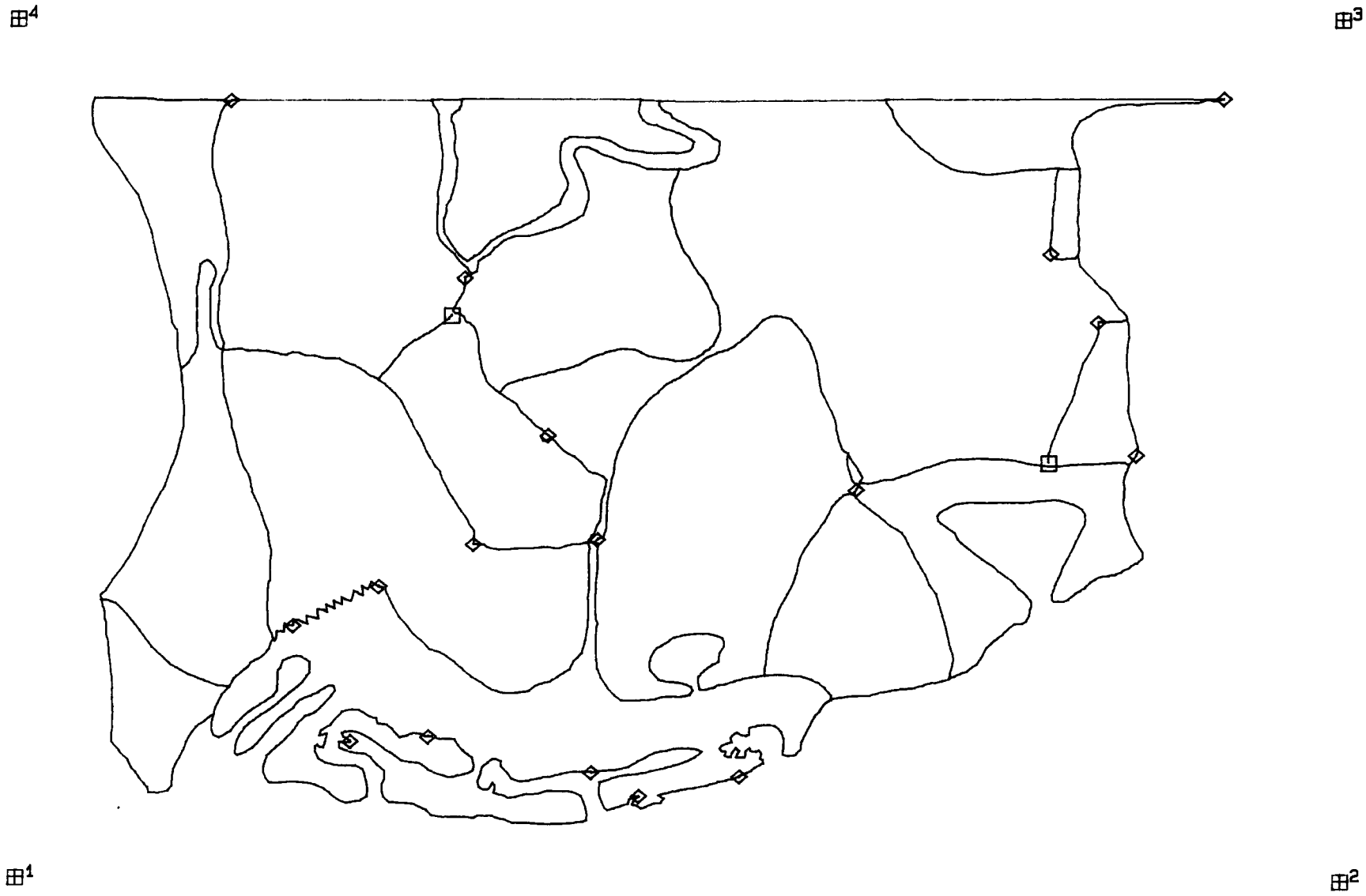
Unecessary lines are 'weeded' out using a set linear tolerance.No lines with a length greater than the tolerance can exist. Thus lines lose their sinuosity, but save on storage space.



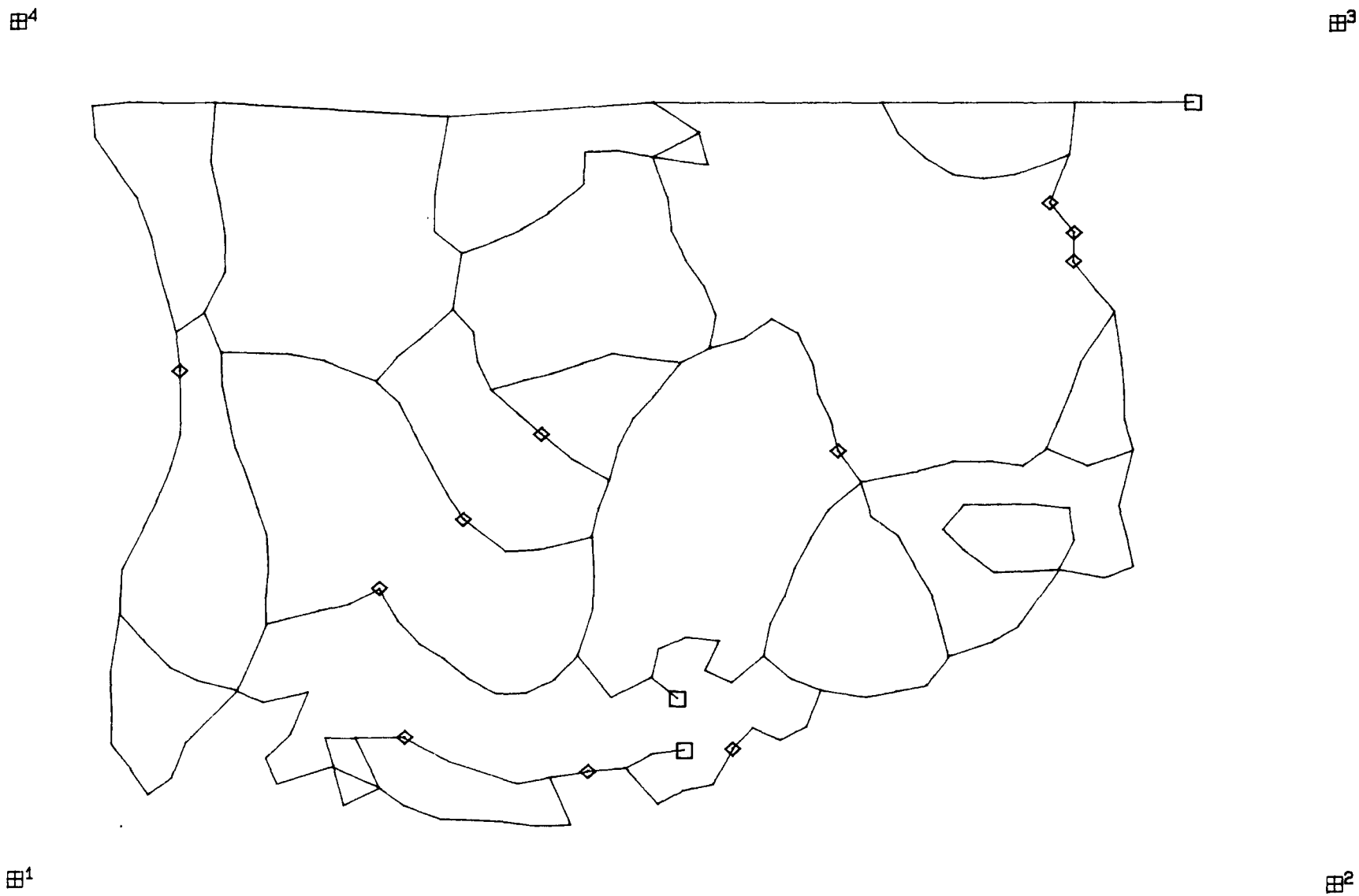
5.13 Original coverage : digitised with both fuzzy and dangle tol = 5 cov units



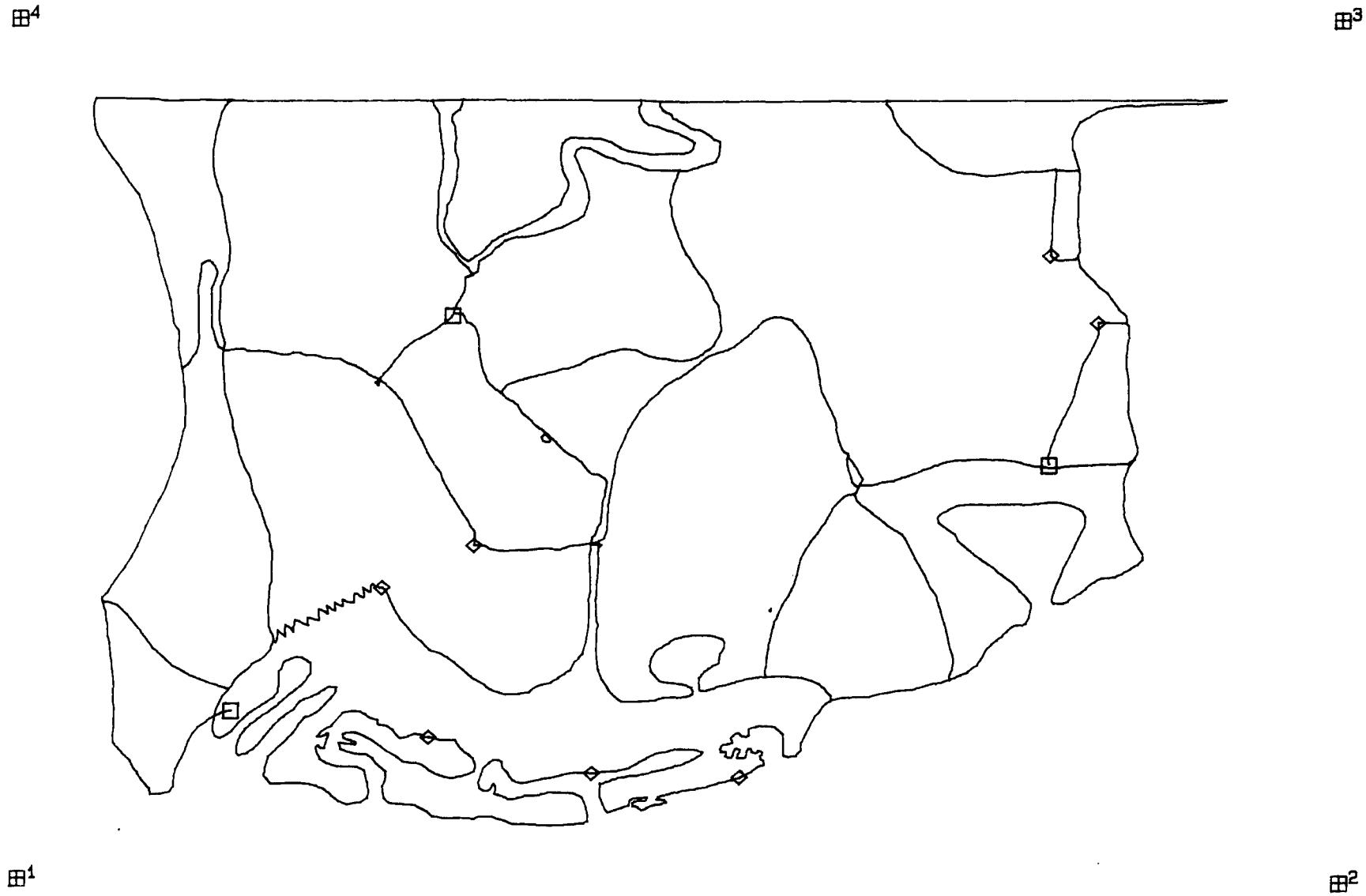
5.14 Fuzzy tol = 50 coverage units : dangle tol = 60 coverage units



5.15 Mnode tol = 30 cov units : fuzzy tol = 5 cov units : dangle tol = 30 cov units



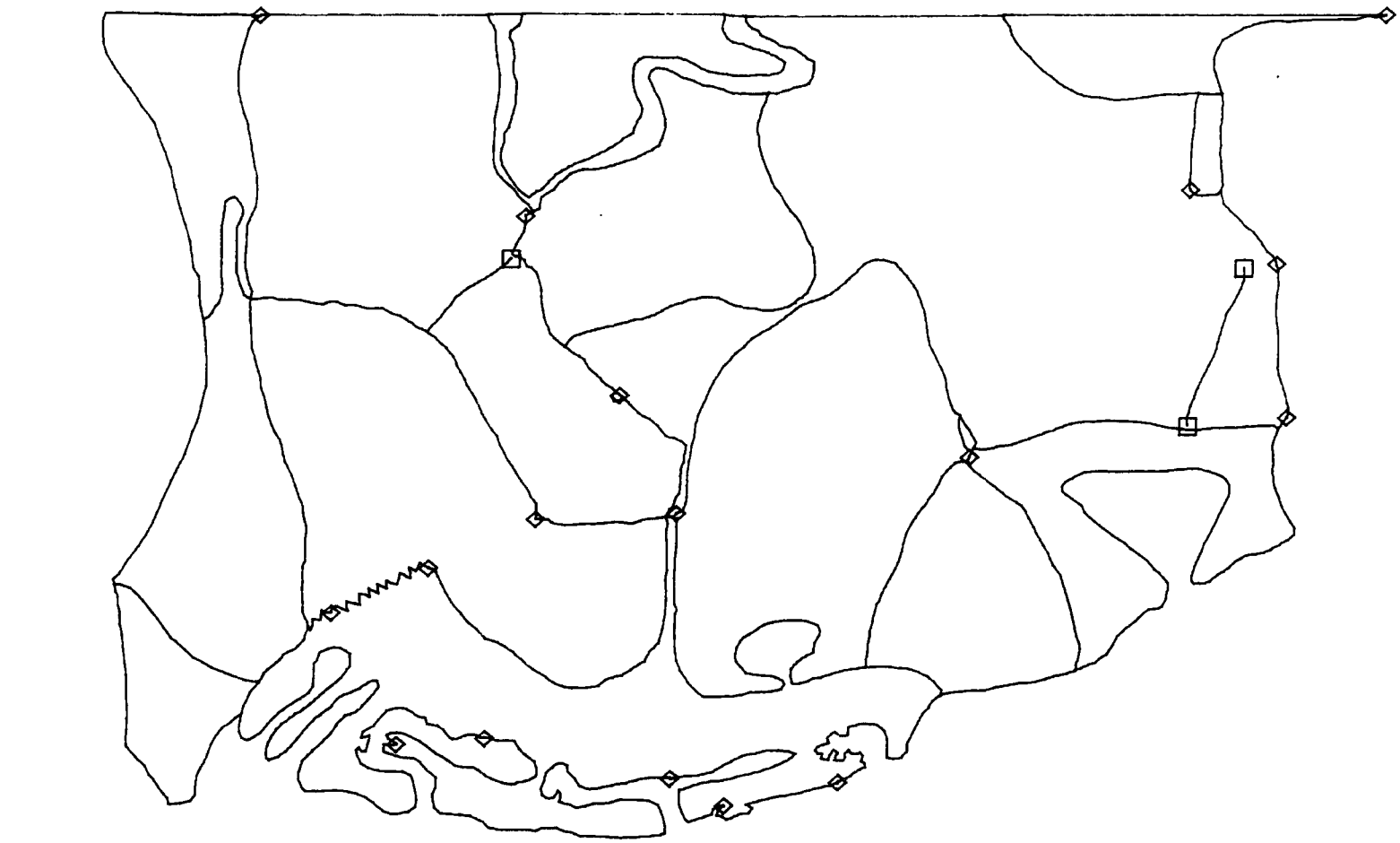
5.16 Fuzzy tol = 50 cov units : dangle tol = 30 cov units



5.17 Mnode = 30 coverage units

5.18 Mnode = 50 coverage units

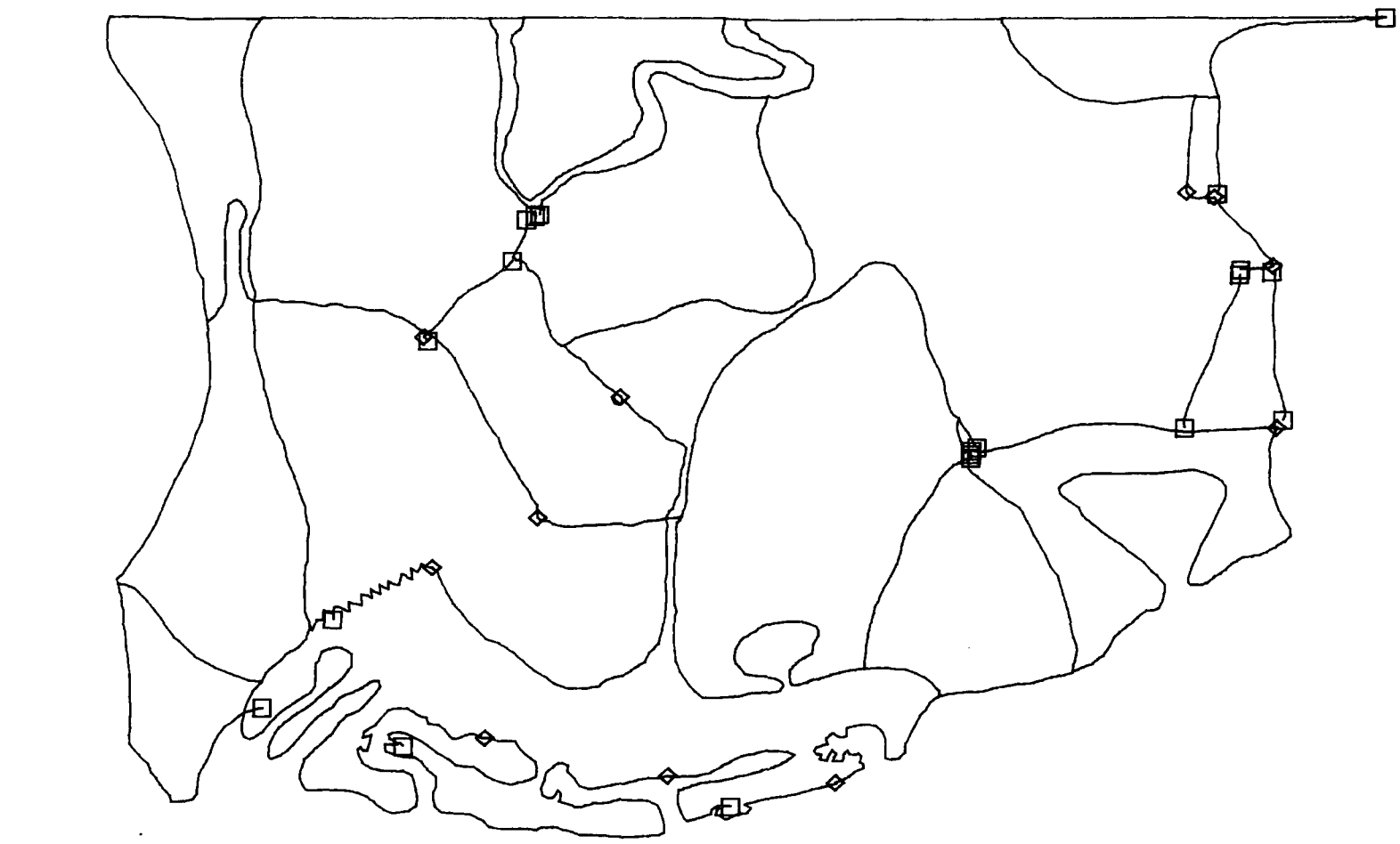
5.19 Mnode tol = 50 cov units : fuzzy tol = 5 cov units : dangle tol = 60 cov units



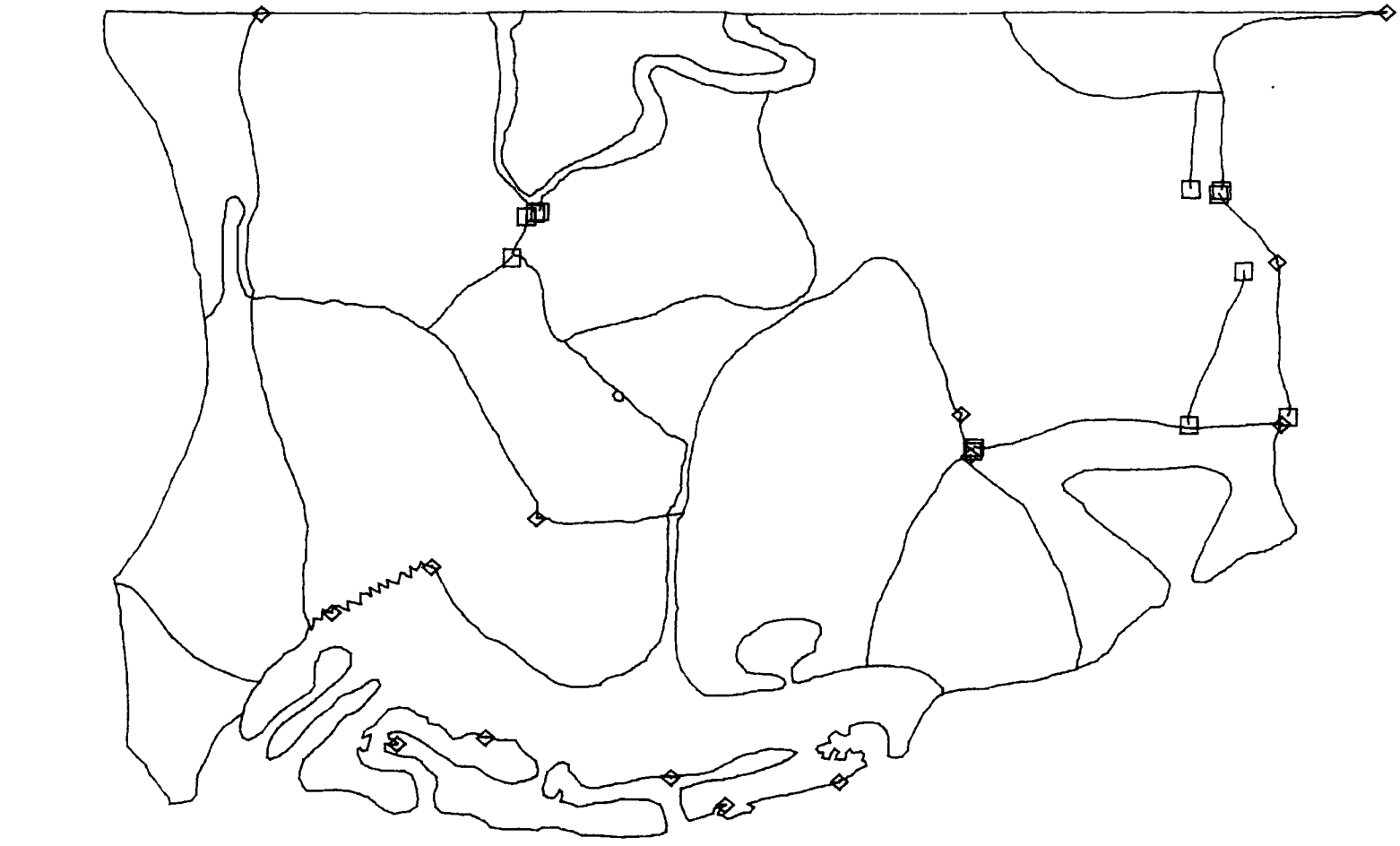
5.20 Mnode tol = 30 cov units : fuzzy tol = 5 cov units : dangle tol = 60 cov units

⊞[1] ⊞[2]

5.21 Fuzzy tol = 8 coverage units : dangle tol = 60 coverage units

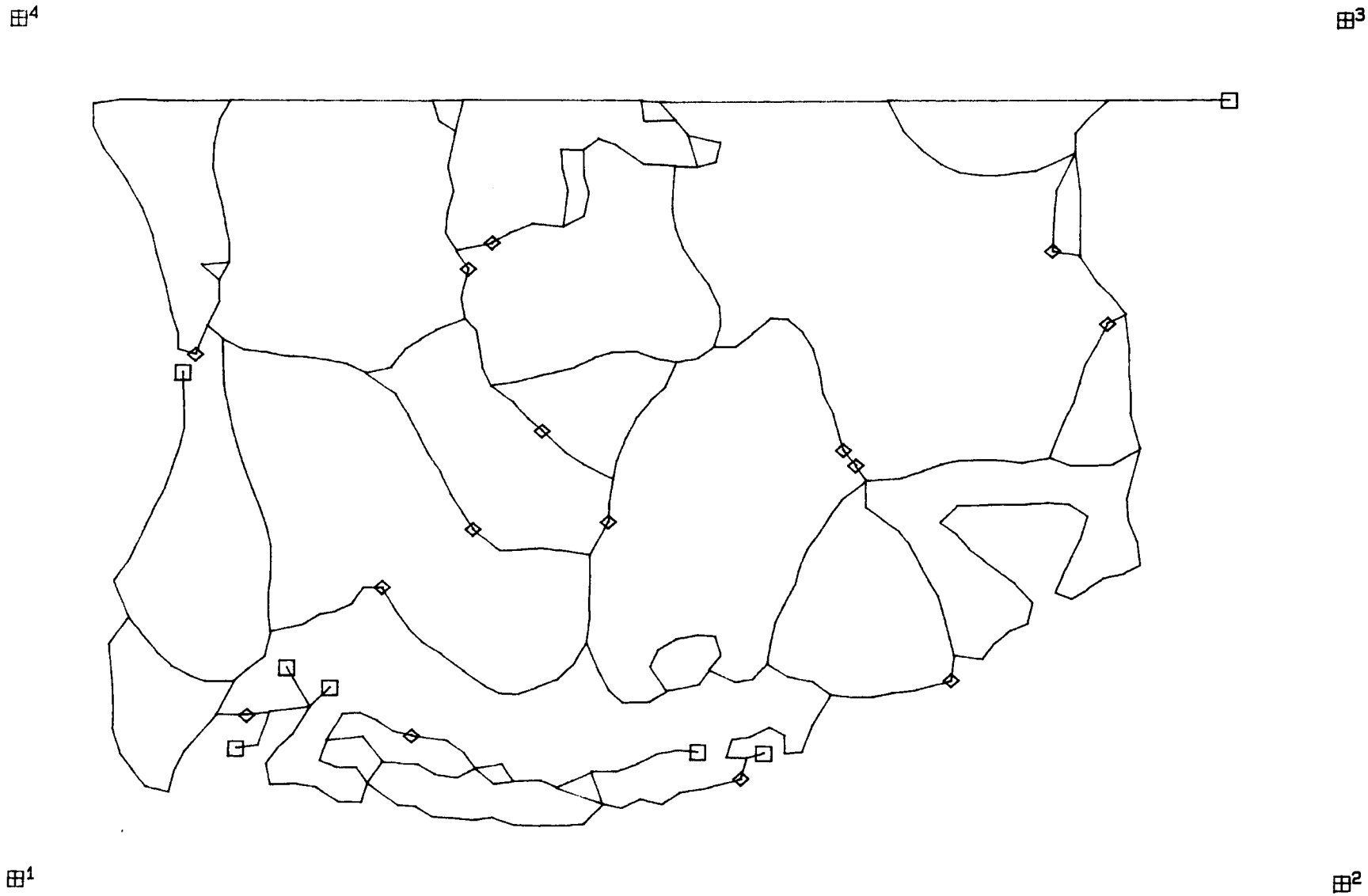


5.22 Mnode tol = 3 cov units : fuzzy tol = 5 cov units : dangle tol = 2 cov units



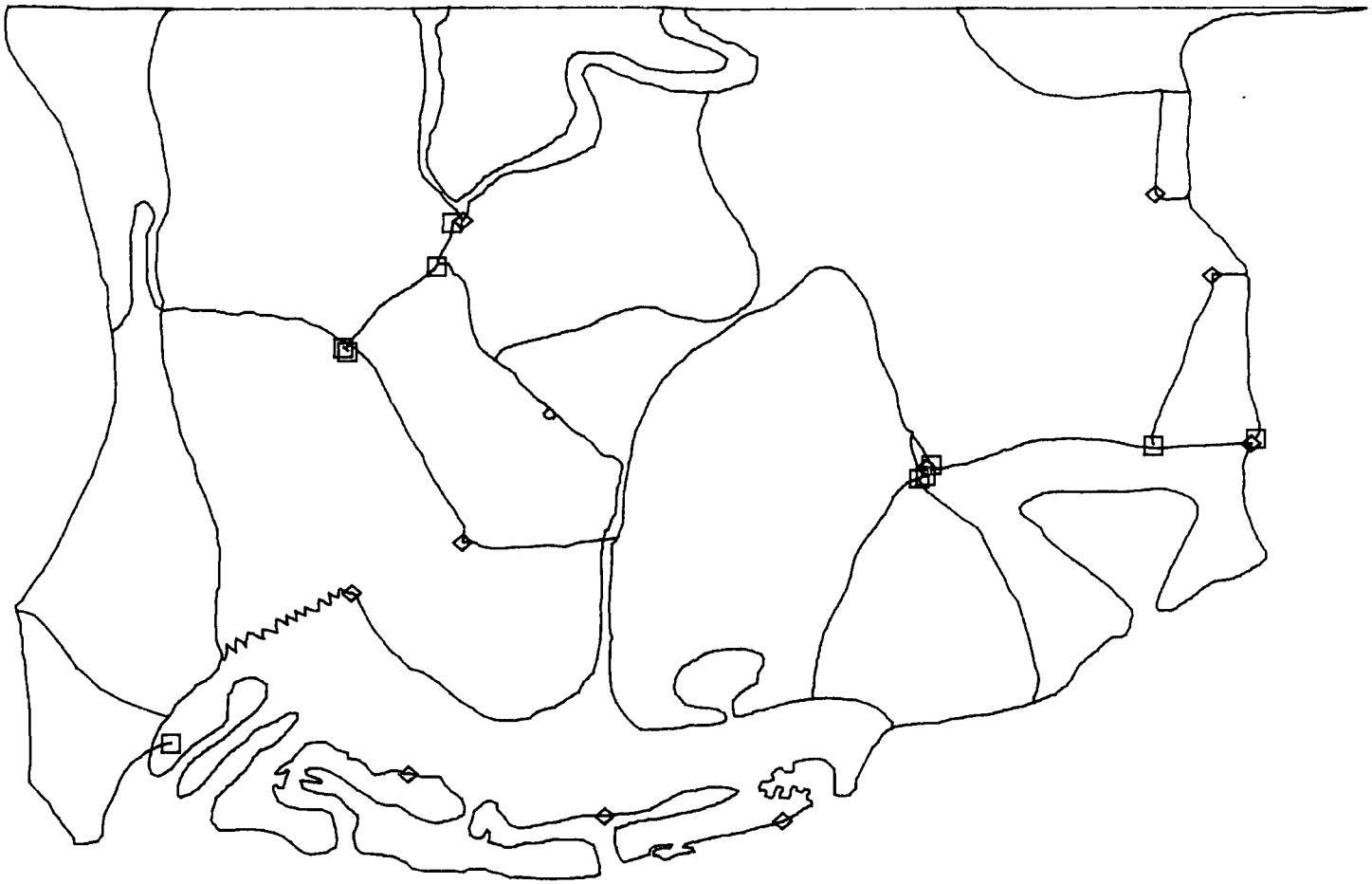
5.23 Mnode tol = 5 cov units : fuzzy tol = 5 cov units : dangle tol = 60 cov units

5.24 Mnode tol = 50 cov units : fuzzy tol = 5 cov units : dangle tol = 30 cov units

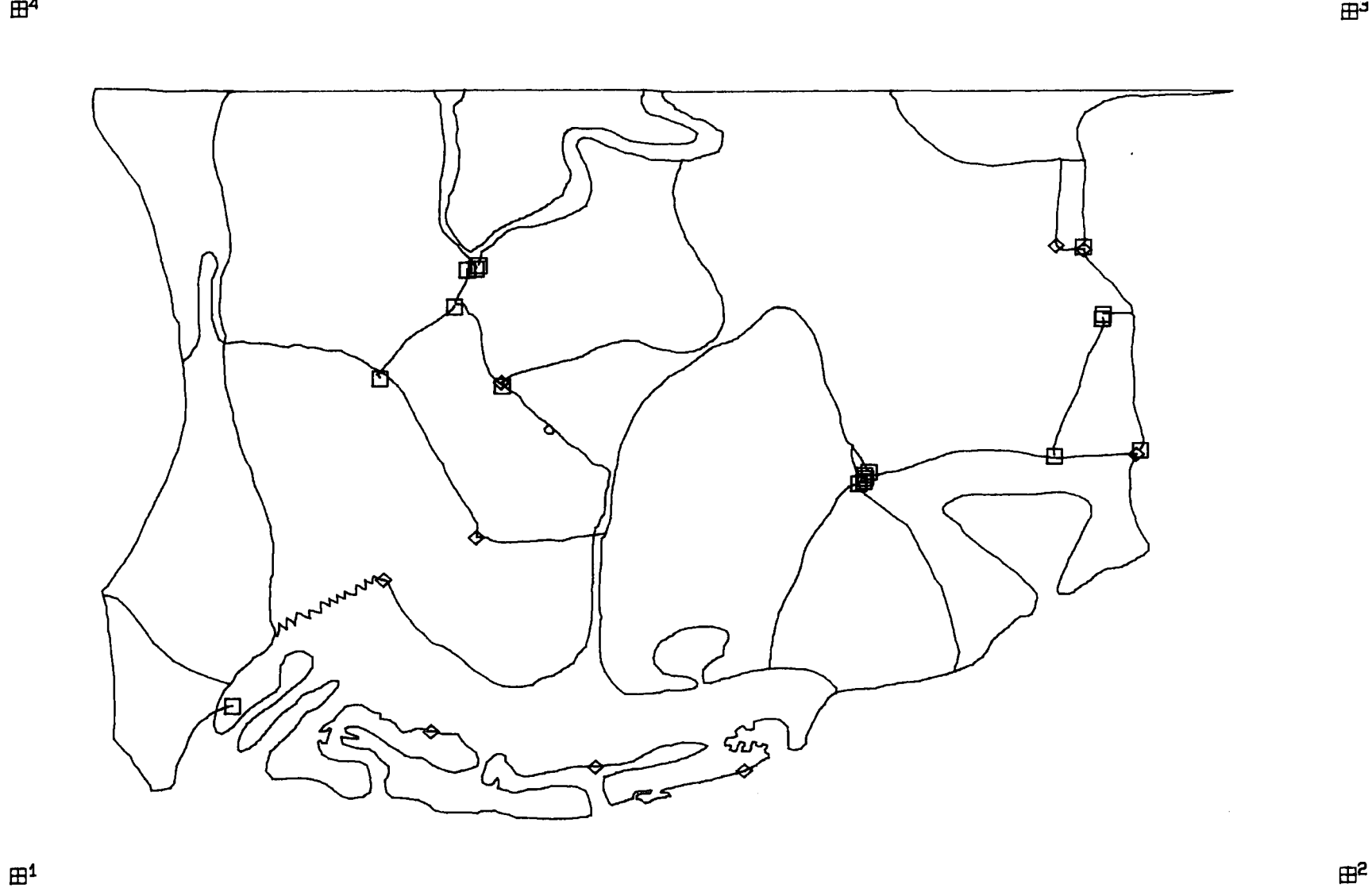


5.25 Fuzzy tol = 30 coverage units ; dangle tol = 8 coverage units

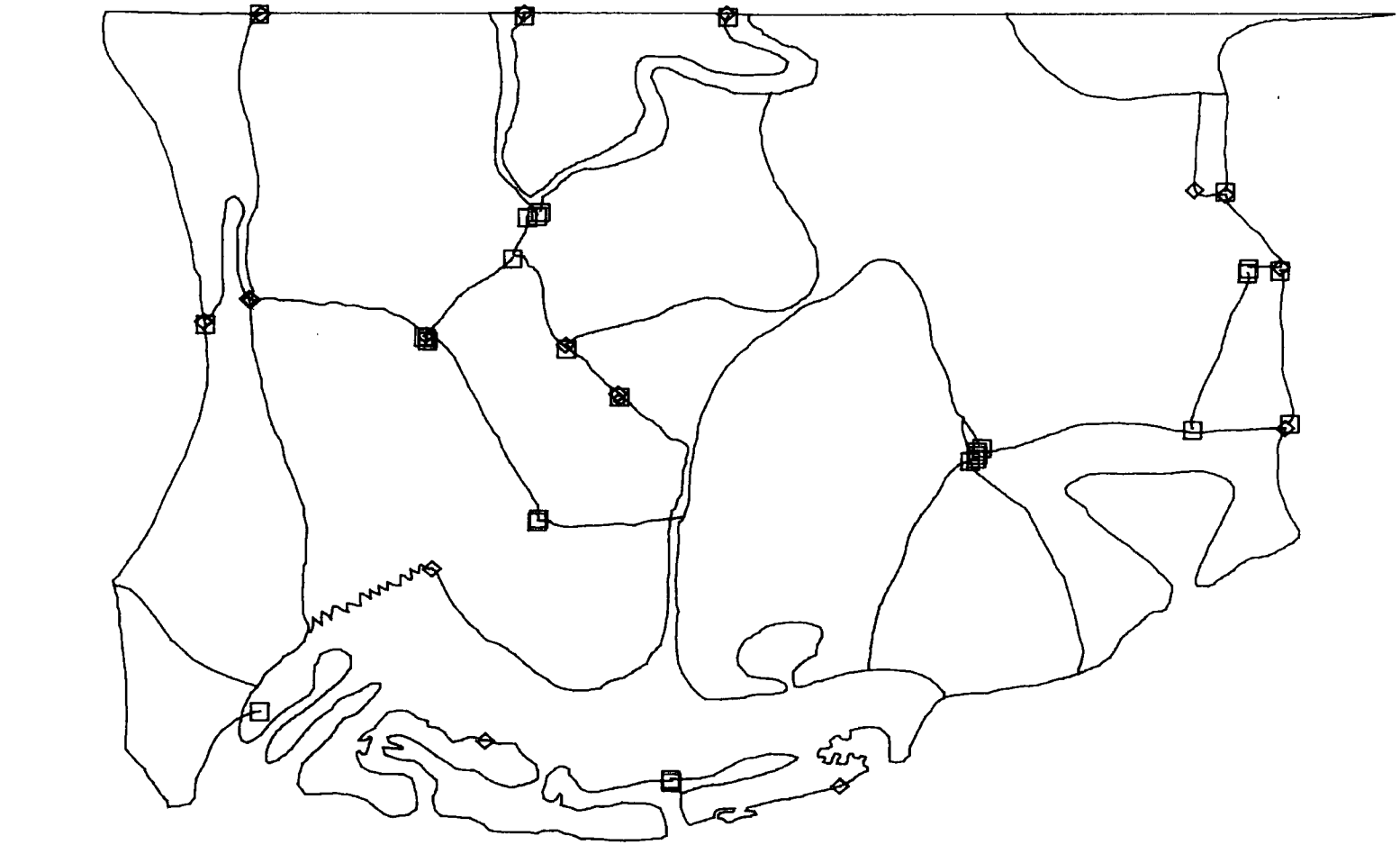
5.26 Mnode = 20 coverage units



5.27 Mnode = 8 coverage units

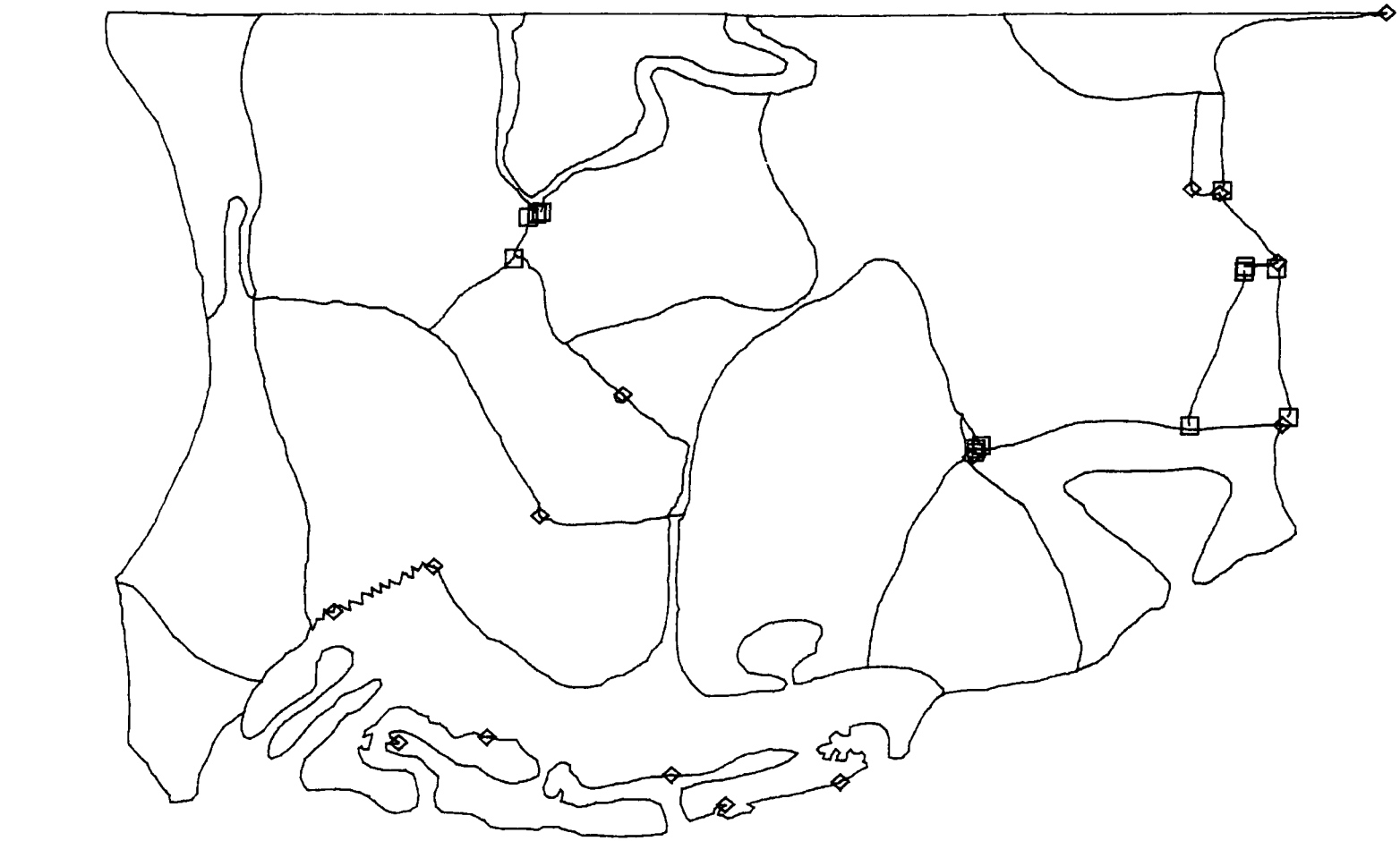


5.28 Mnode = 5 coverage units

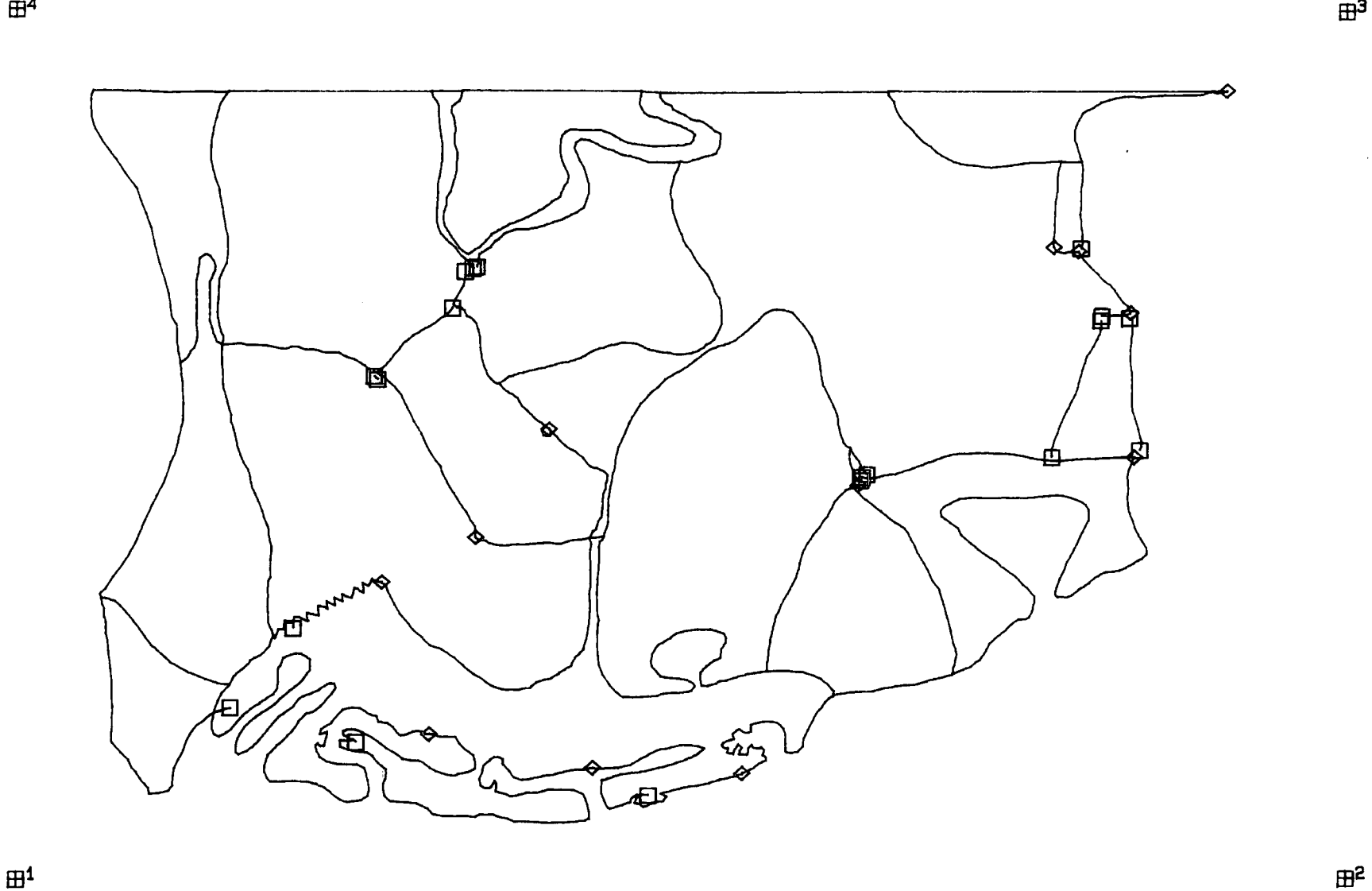


5.29 Mnode = 3 coverage units

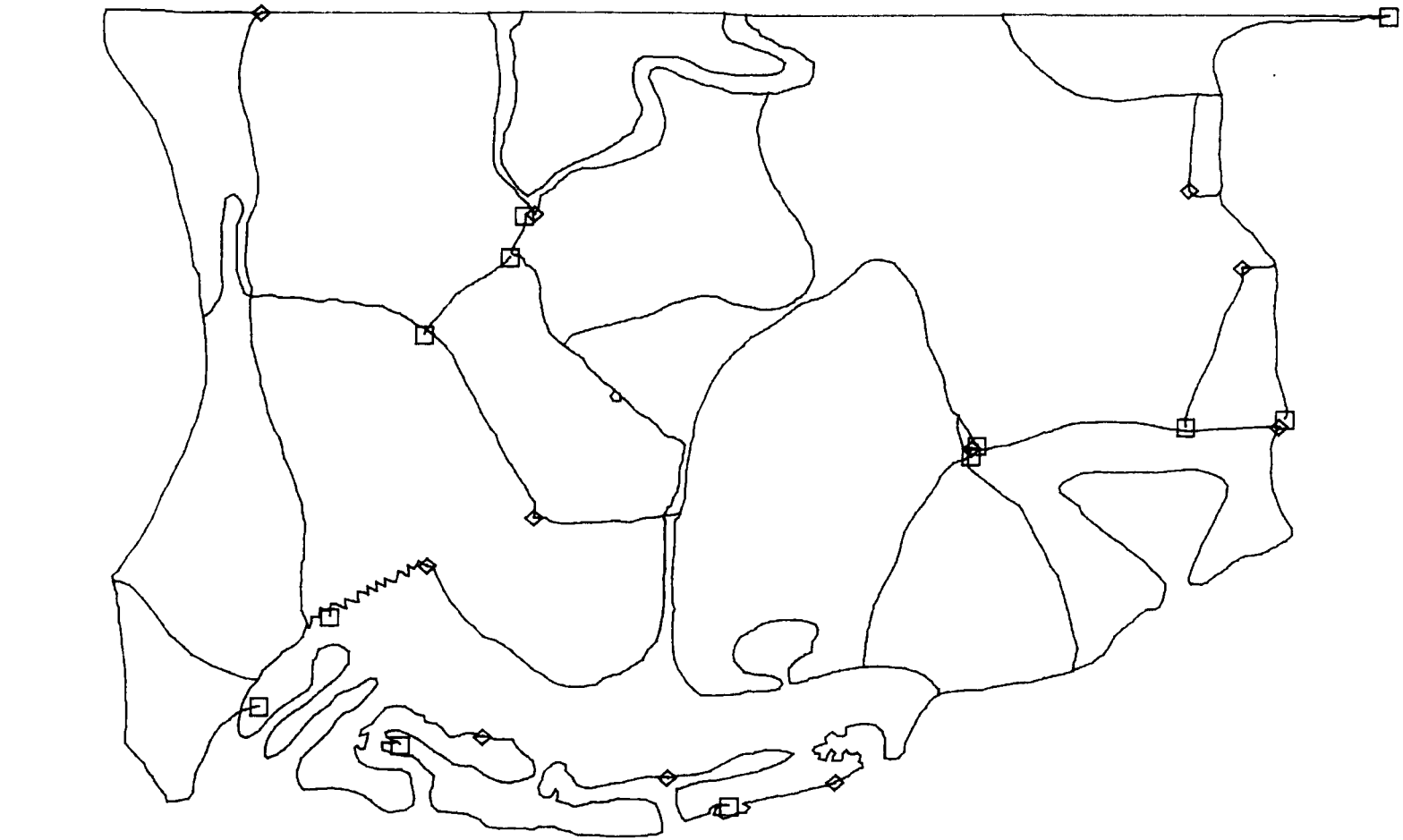
5.30 Mnode tol = 3 cov units : fuzzy tol = 5 cov units : dangle tol = 60 cov units



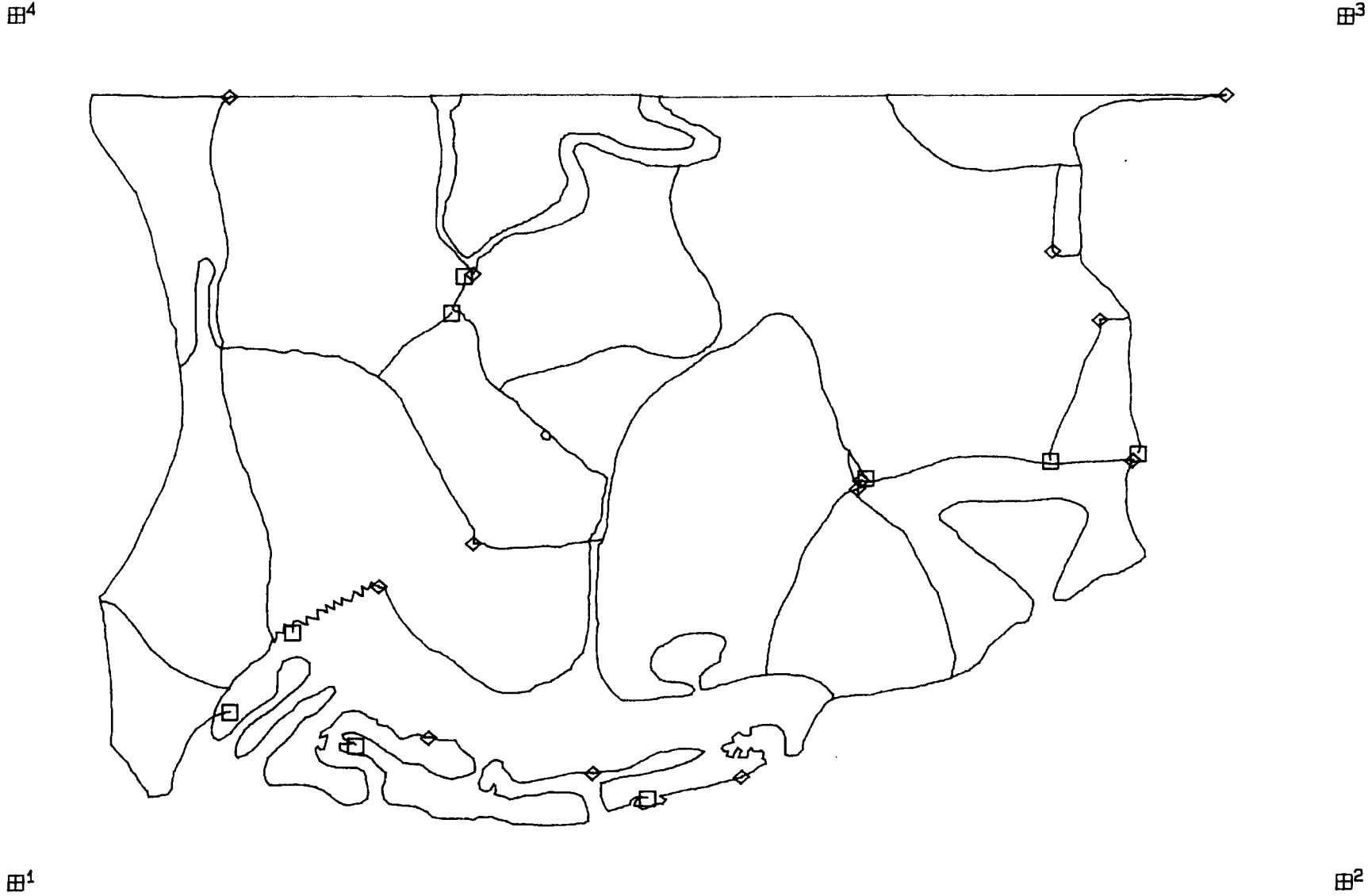
5.31 Mnode tol = 3 cov units : fuzzy tol = 5 cov units : dangle tol = 30 cov units



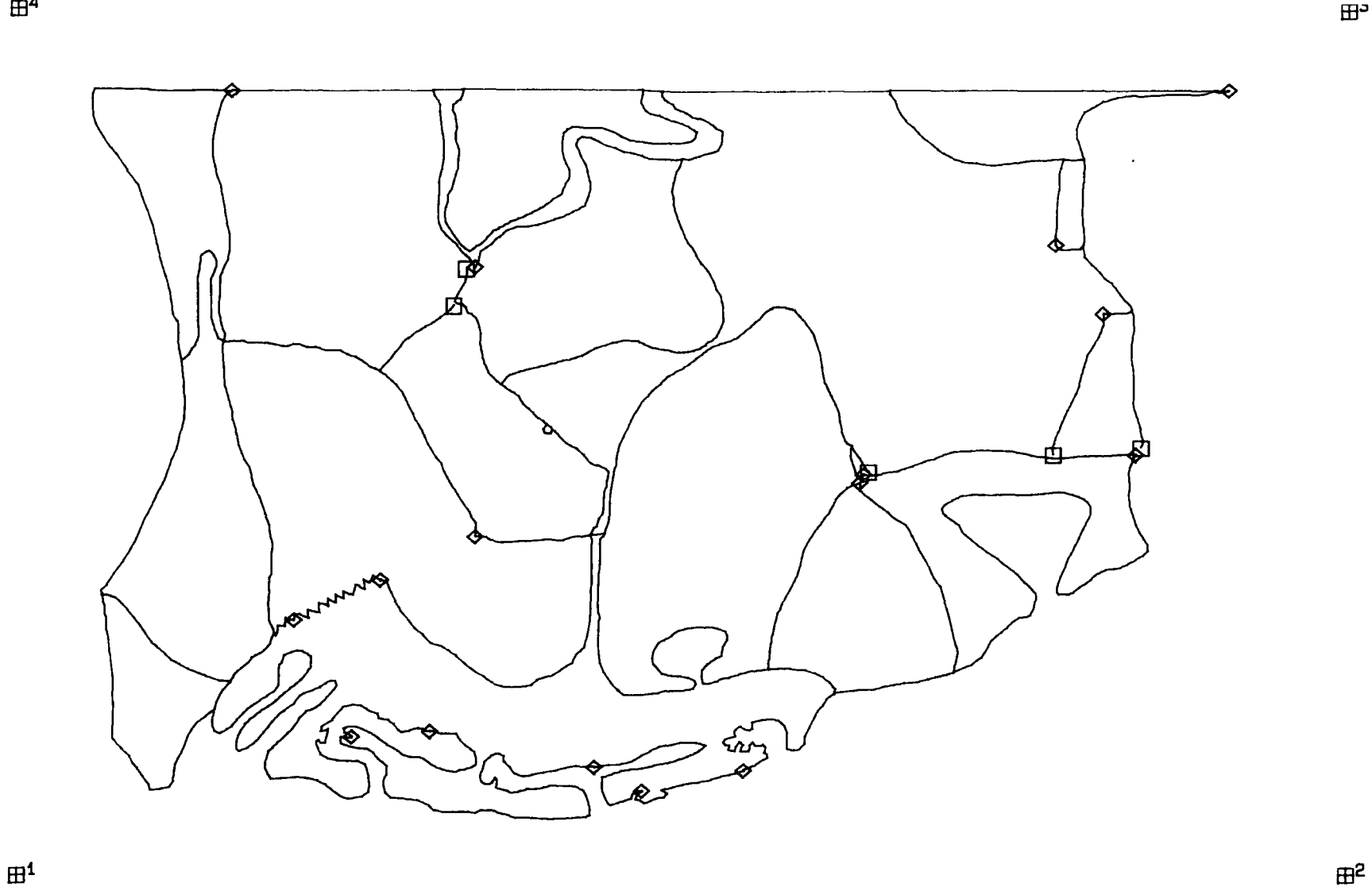
5.32 Mnode tol = 3 cov units : fuzzy tol = 5 cov units : dangle tol = 8 cov units



5.33 Mnode tol = 8 cov units : fuzzy tol = 5 cov units : dangle tol = 2 cov units

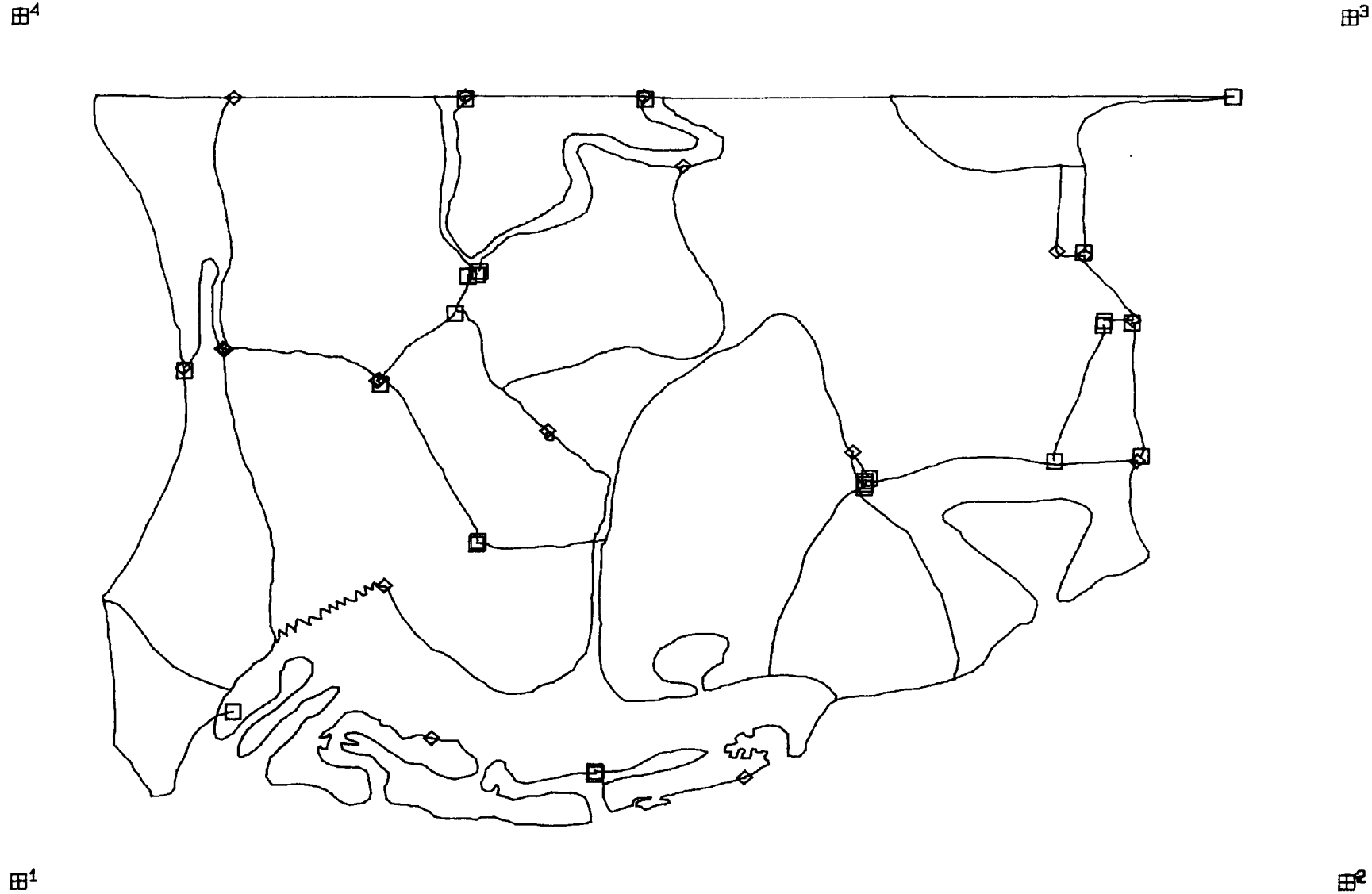


5.34 Mnode tol = 8 cov units : fuzzy tol = 5 cov units : dangle tol = 8 cov units

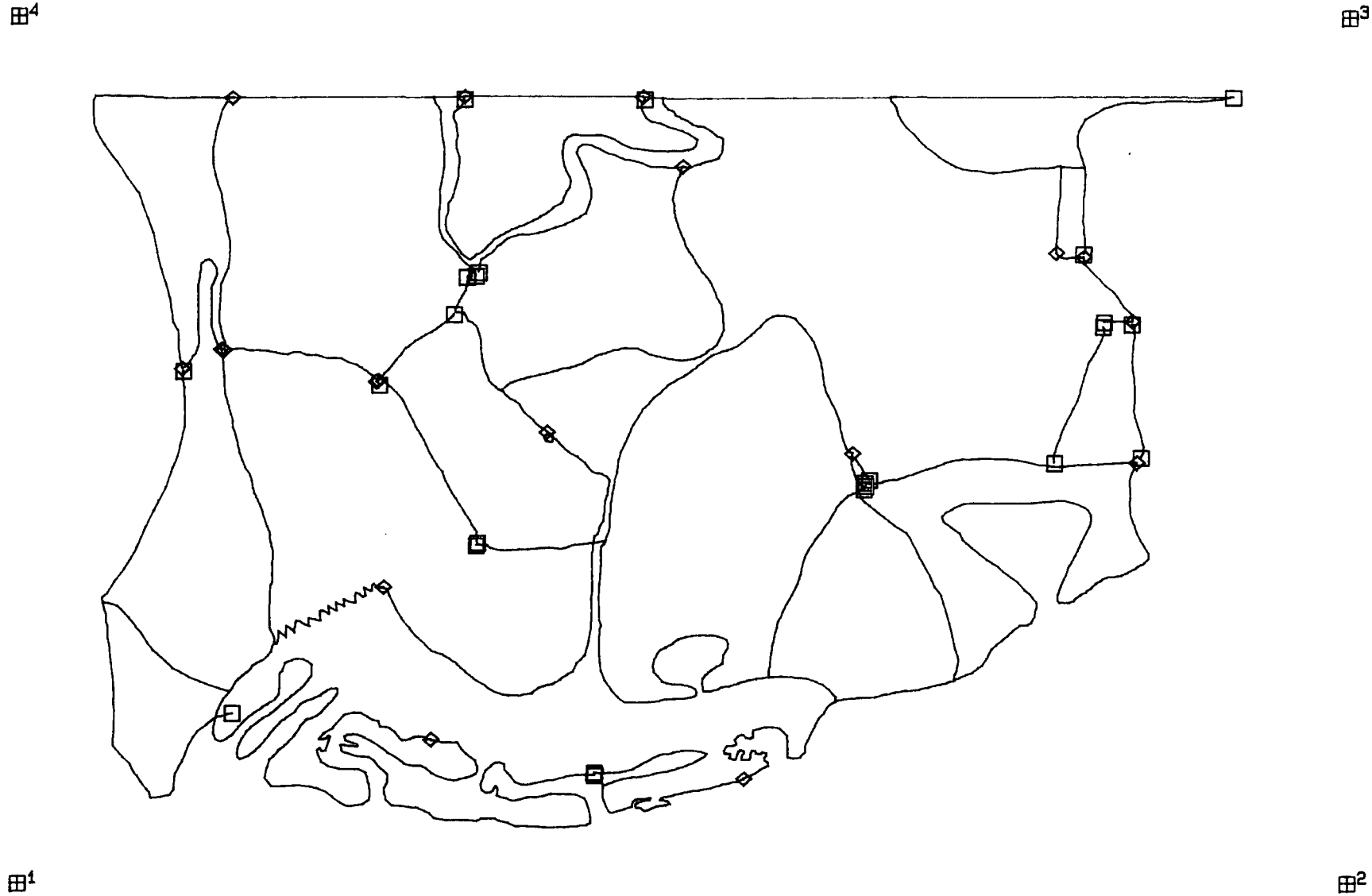


5.35 Mnode tol = 8 cov units : fuzzy tol = 5 cov units : dangle tol = 30 cov units

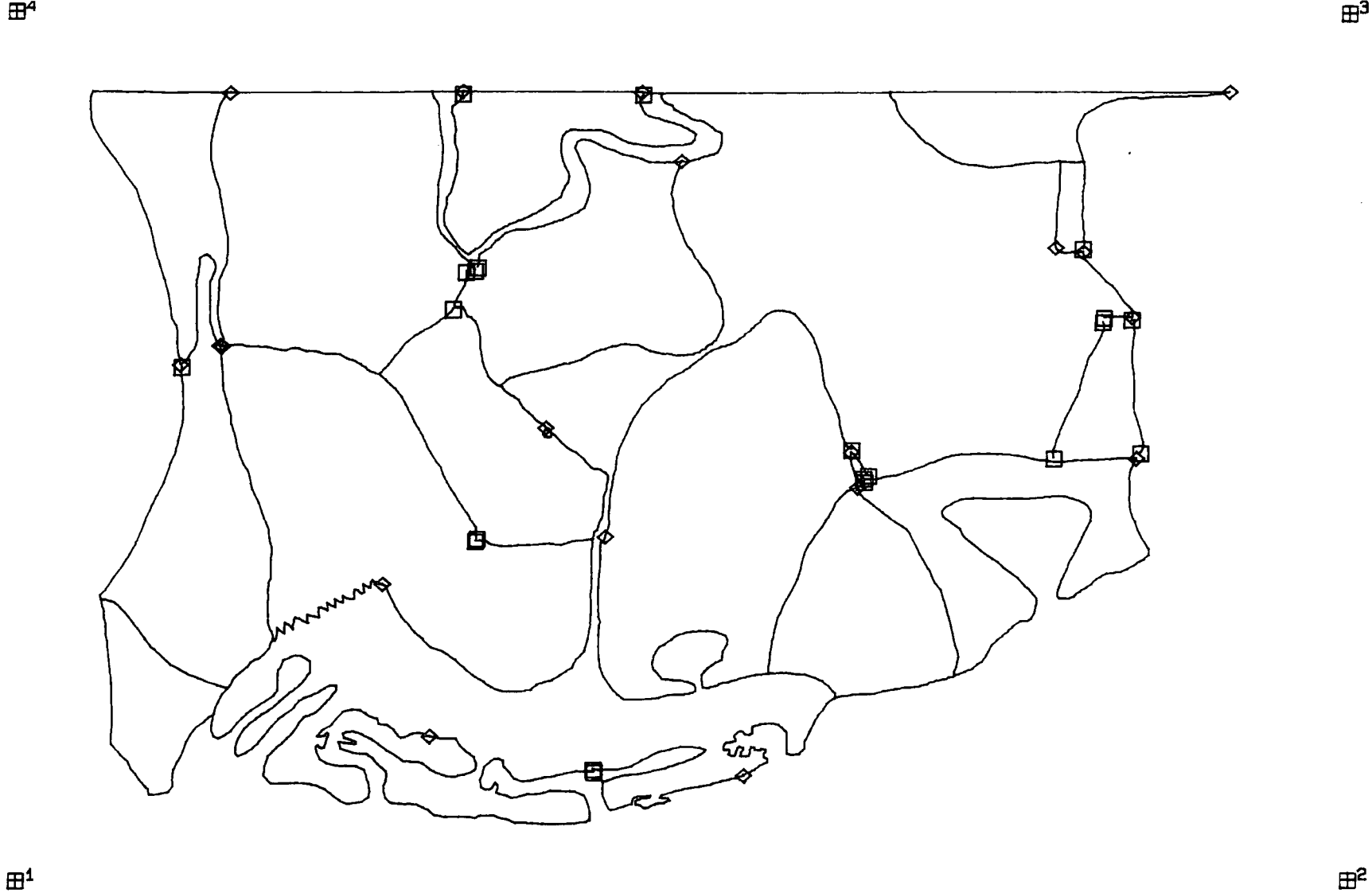
5.36 Mnode tol = 8 cov units : fuzzy tol = 5 cov units : dangle tol = 60 cov units



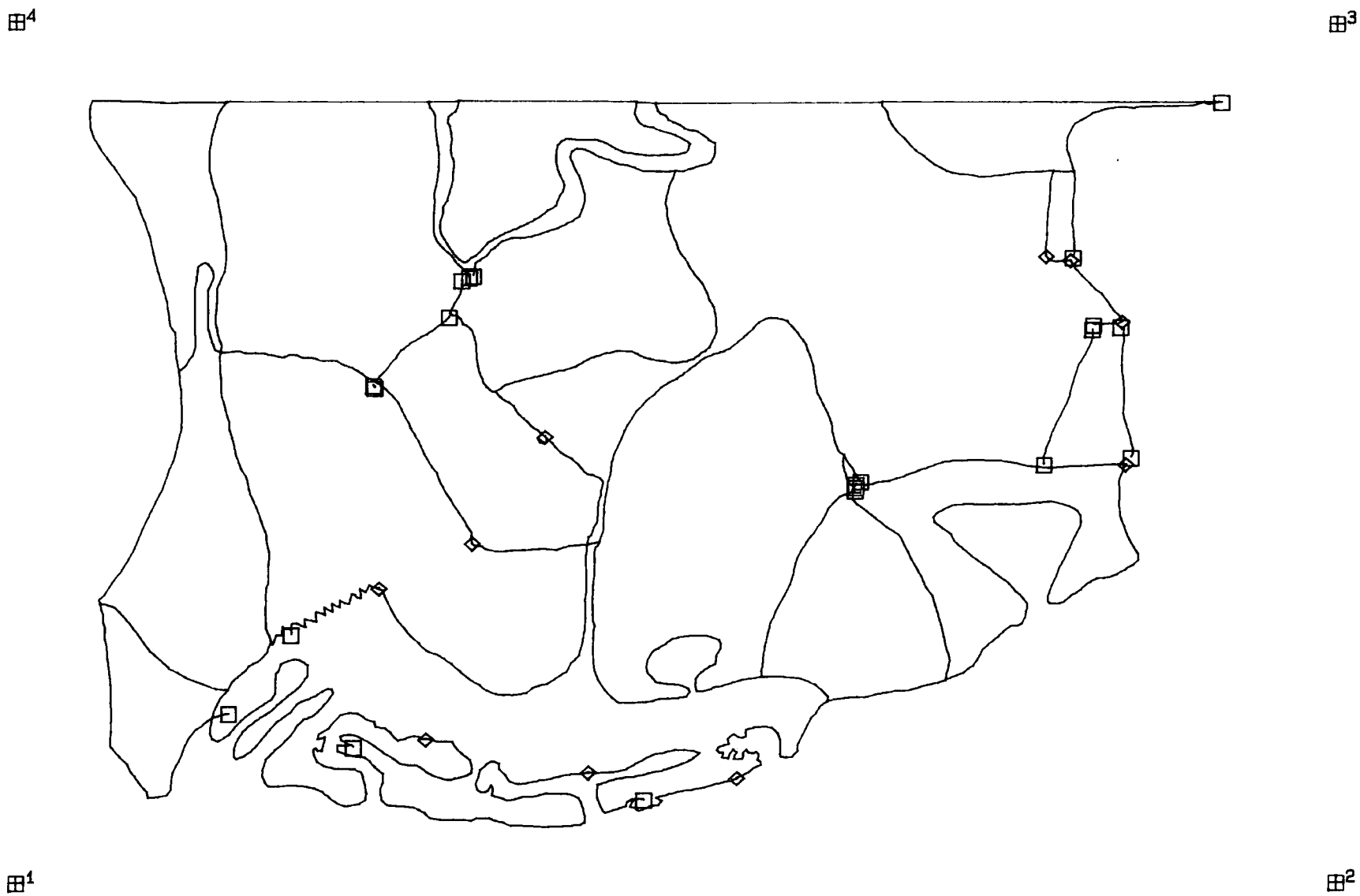
5.37 Fuzzy tolerance = 3 coverage units



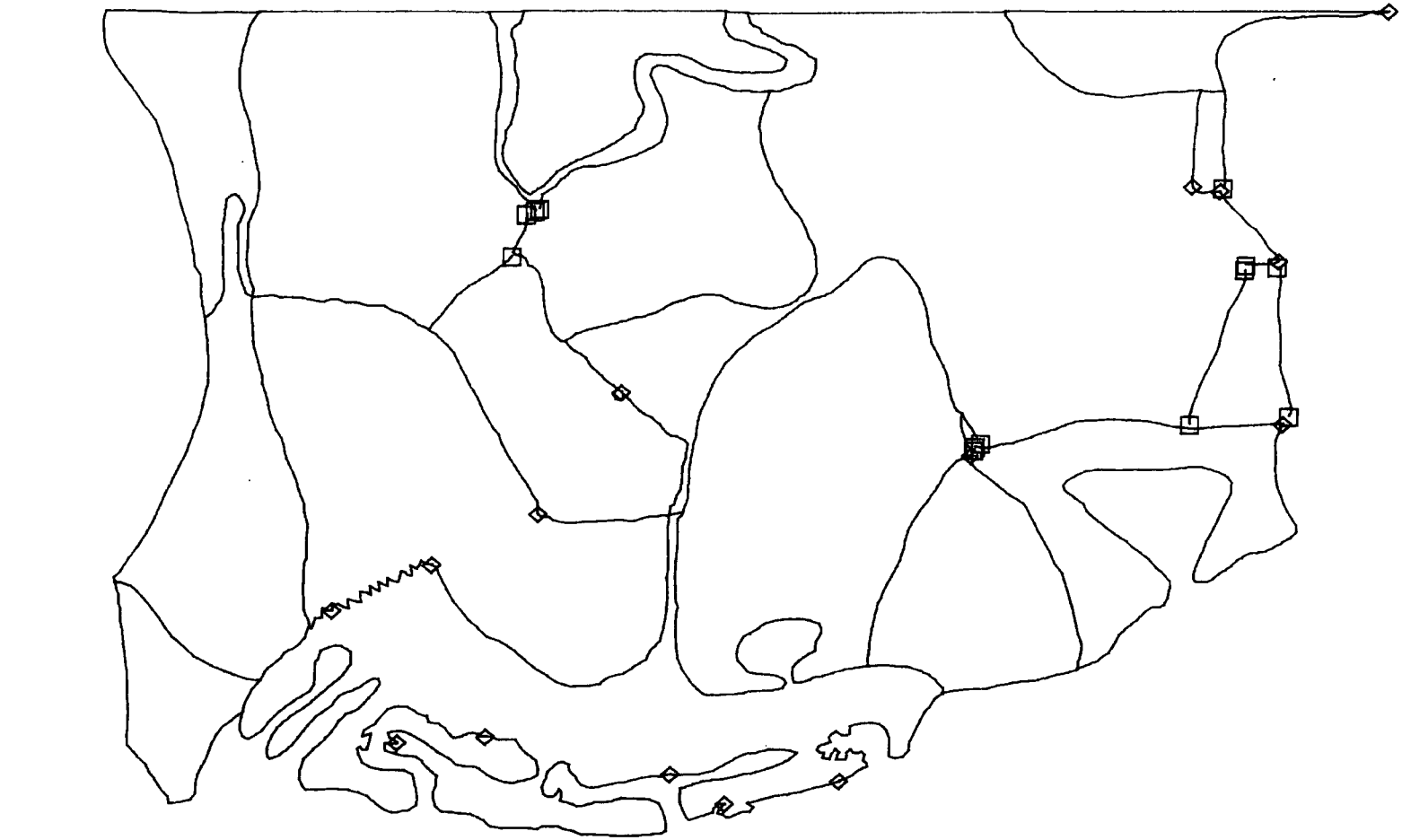
5.38 Fuzzy tol = 3 coverage units : dangle tol = 2 coverage units



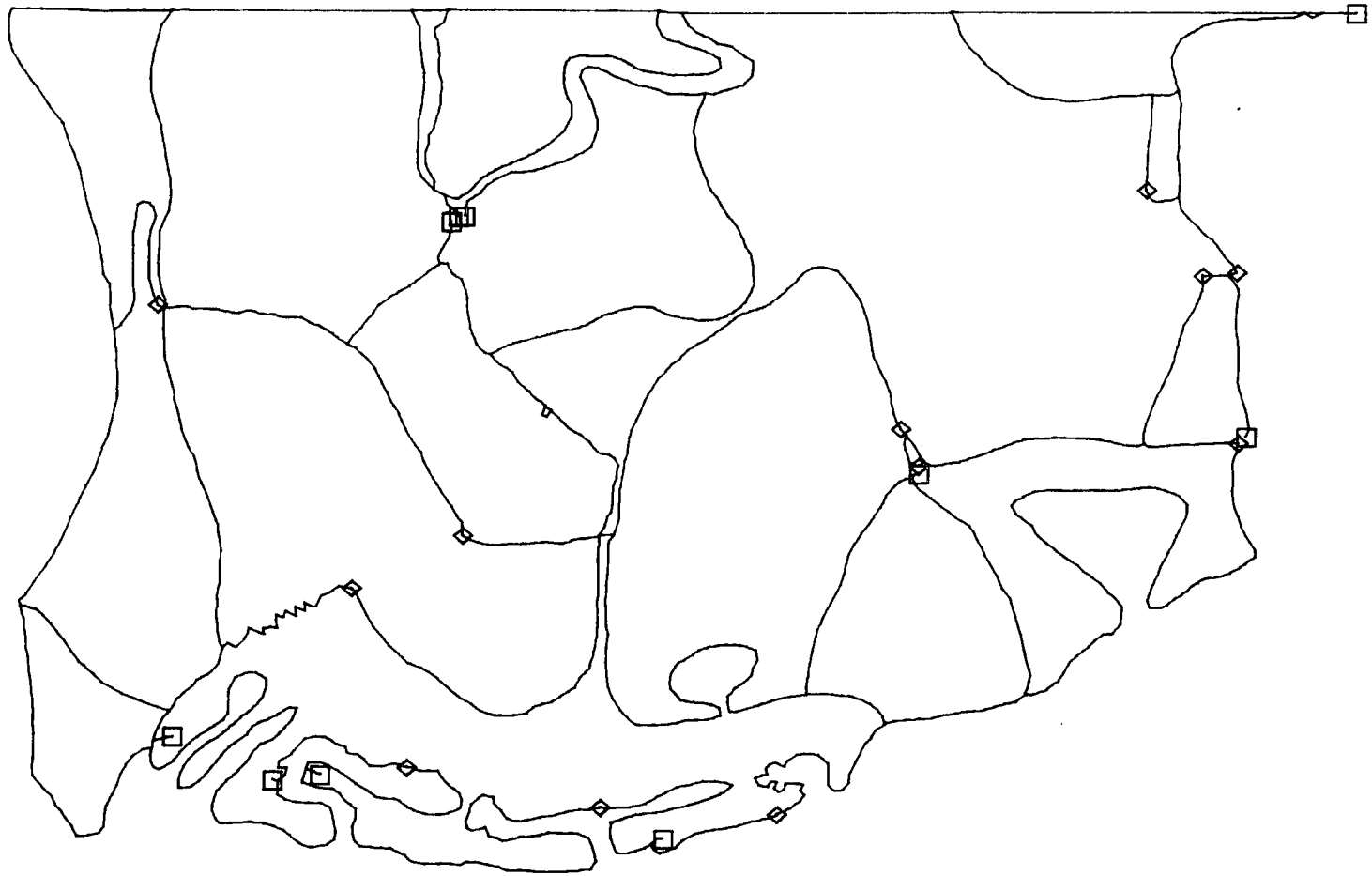
5.39 Fuzzy tol = 3 coverage units : dangle tol = 30 coverage units



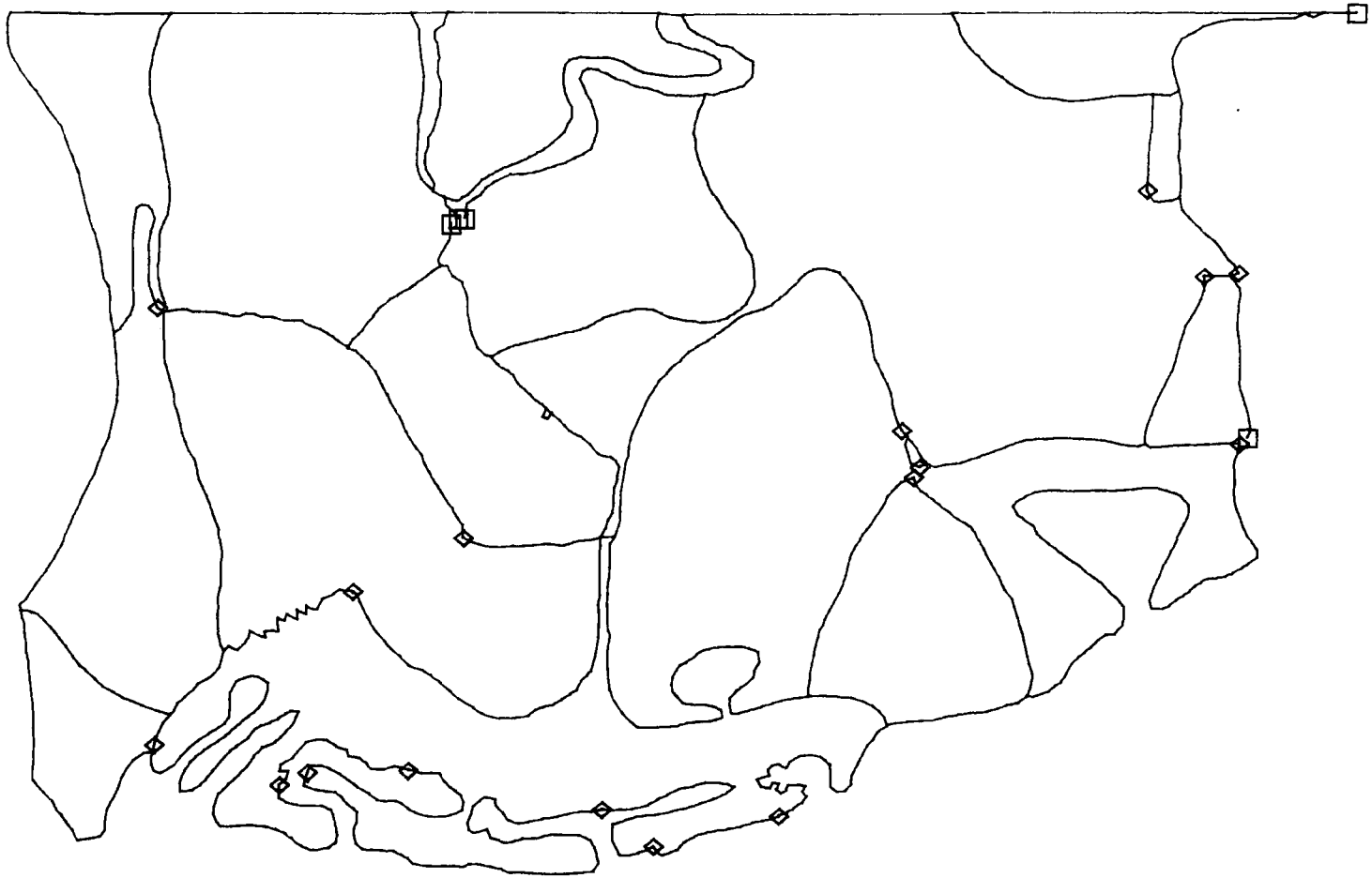
5.40 Fuzzy tol = 5 coverage units : dangle tol = 2 coverage units



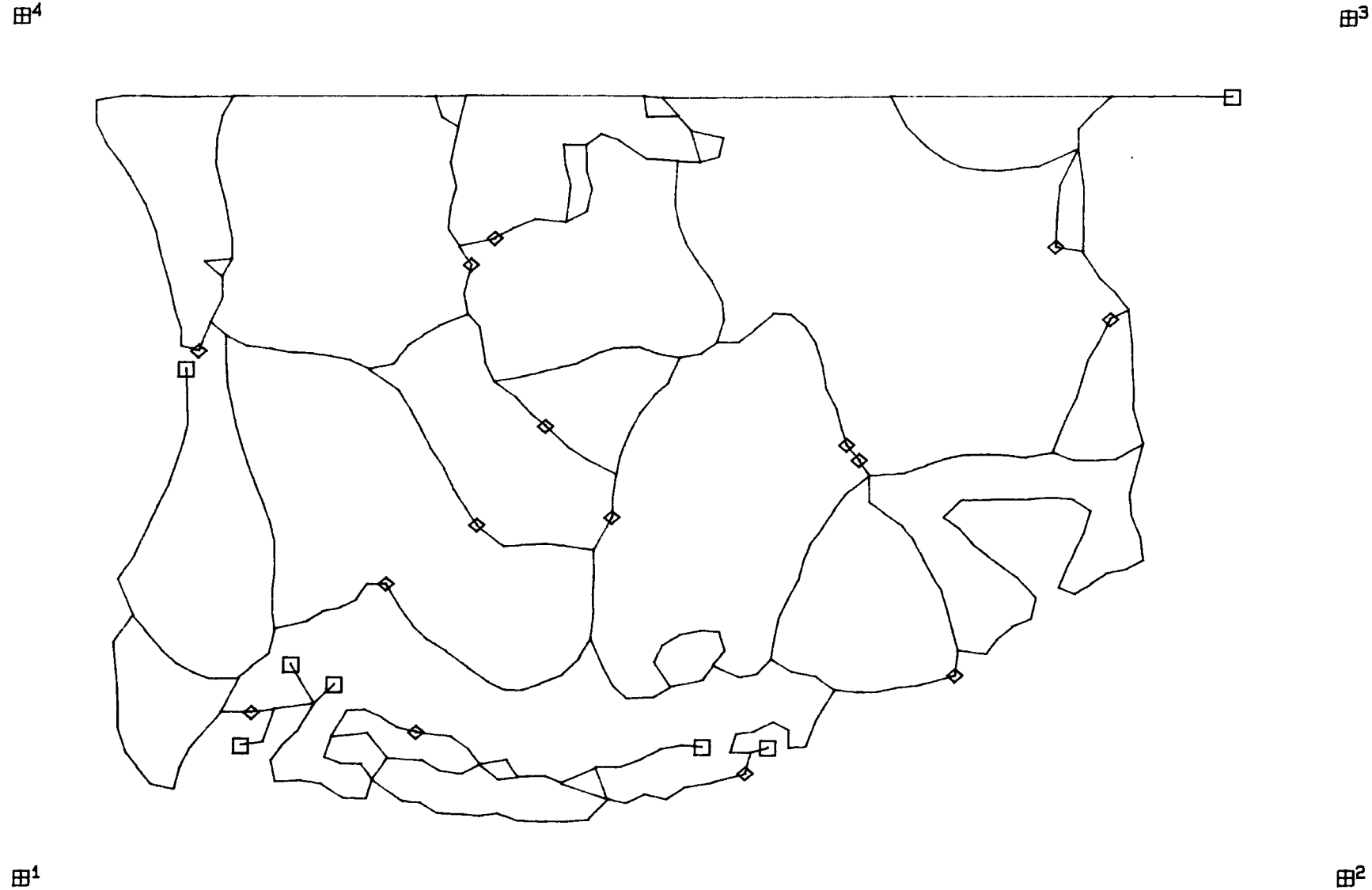
5.41 Fuzzy tol = 5 coverage units : dangle tol = 30 coverage units



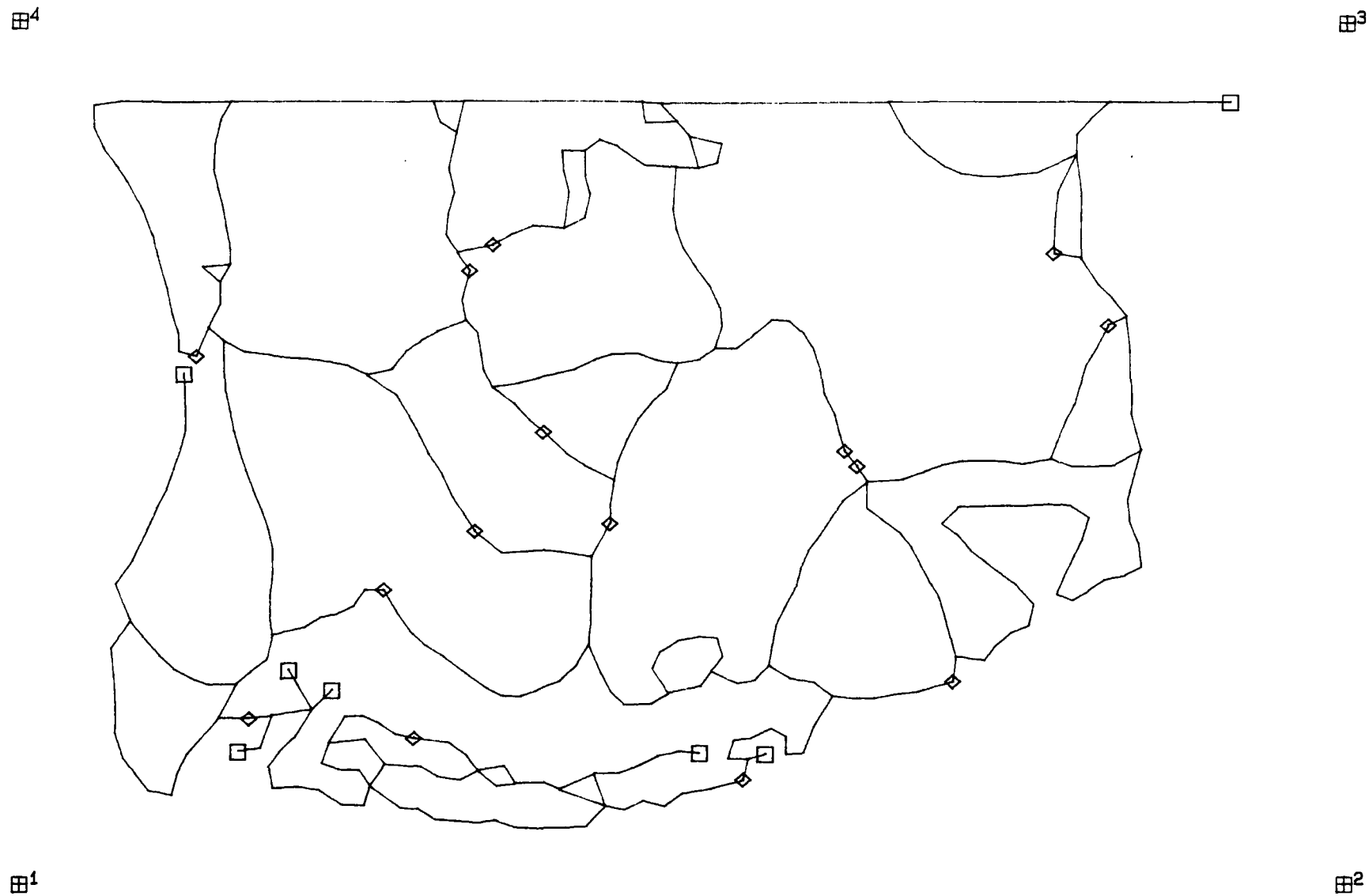
5.42 Fuzzy tol = 8 coverage units : dangle tol = 2 coverage units



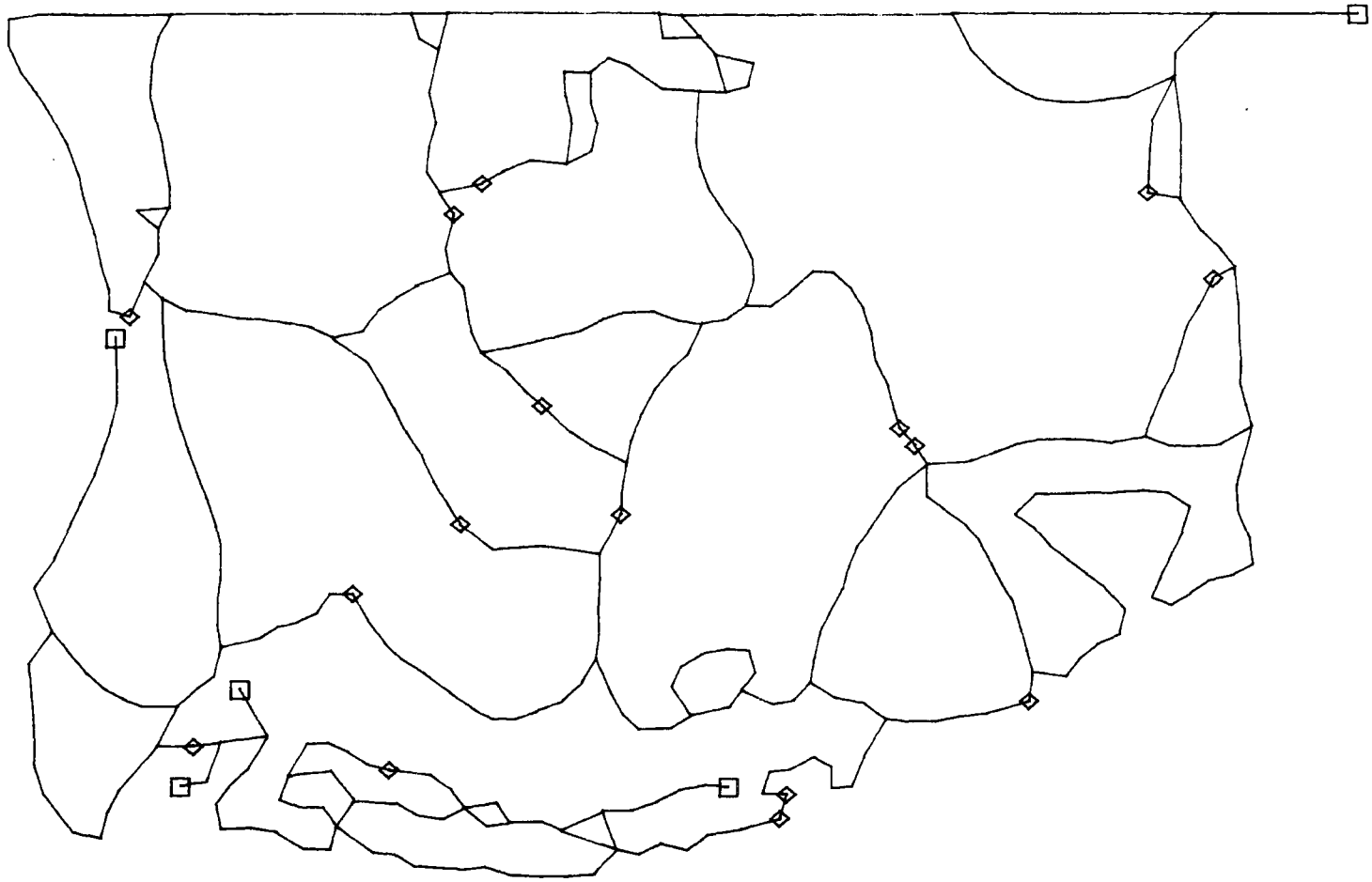
5.43 Fuzzy tol = 8 coverage units : dangle tol = 30 cov units



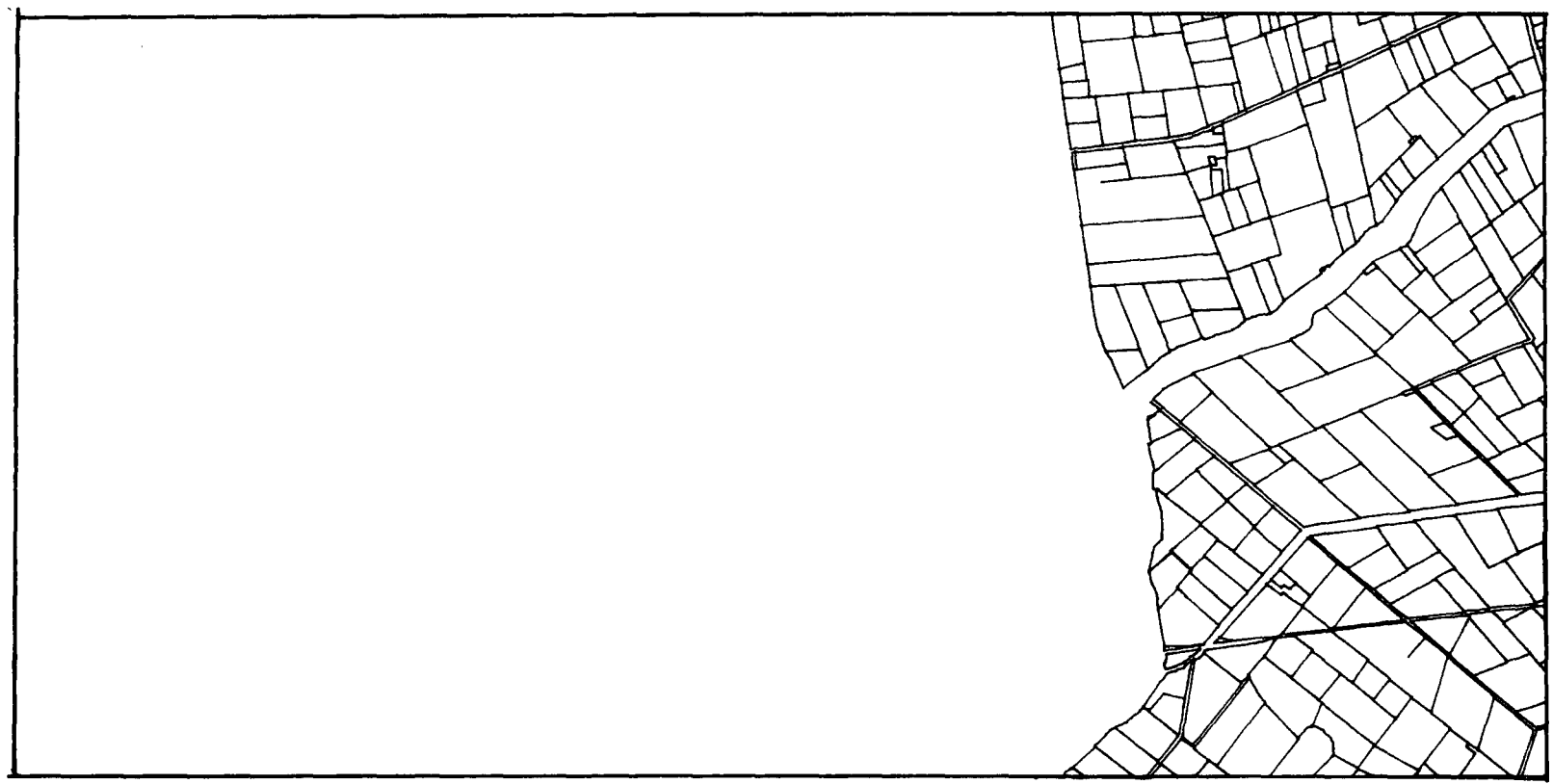
5.44 Fuzzy tol = 30 coverage units : dangle tol = 2 coverage units



5.45 Fuzzy tol = 30 coverage units : dangle tol = 30 coverage units

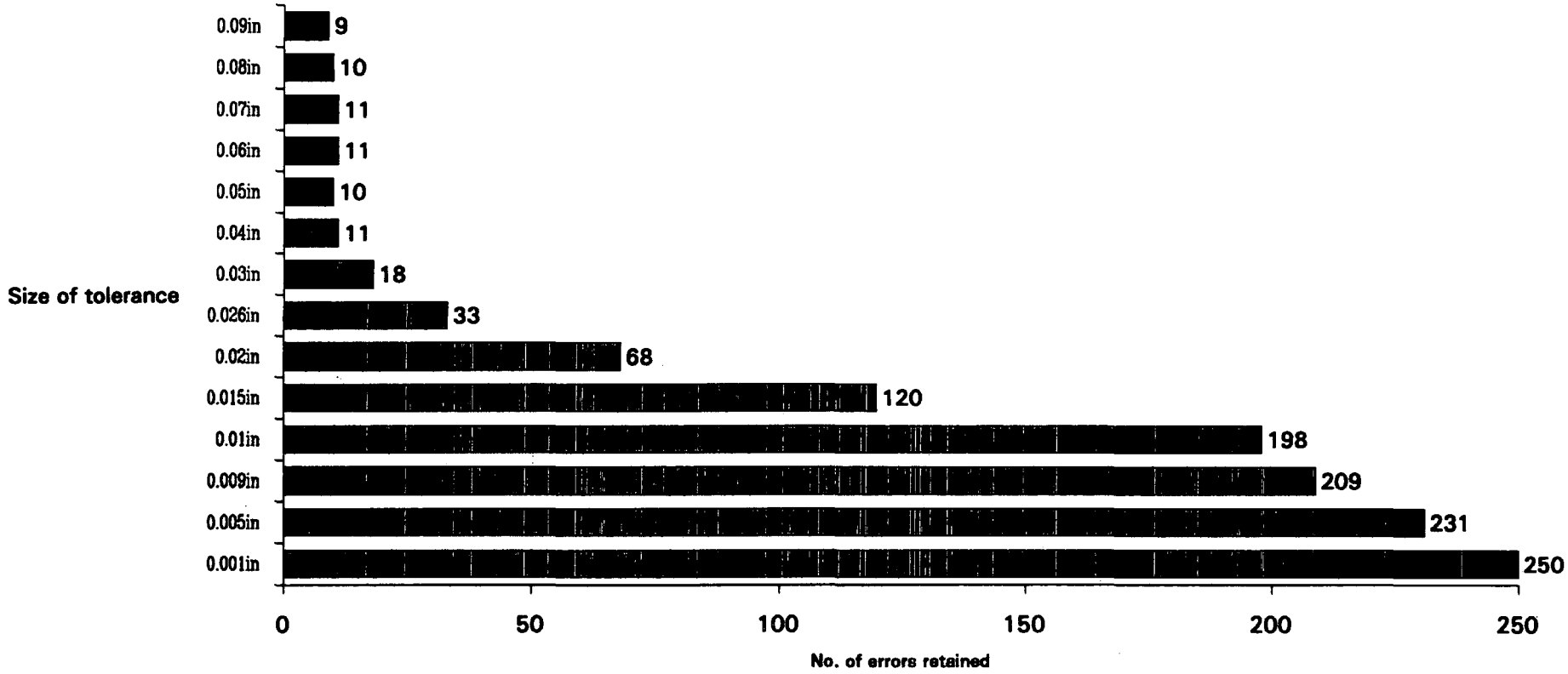


5.46 Fuzzy tol = 30 coverage units : dangle tol = 60 coverage units

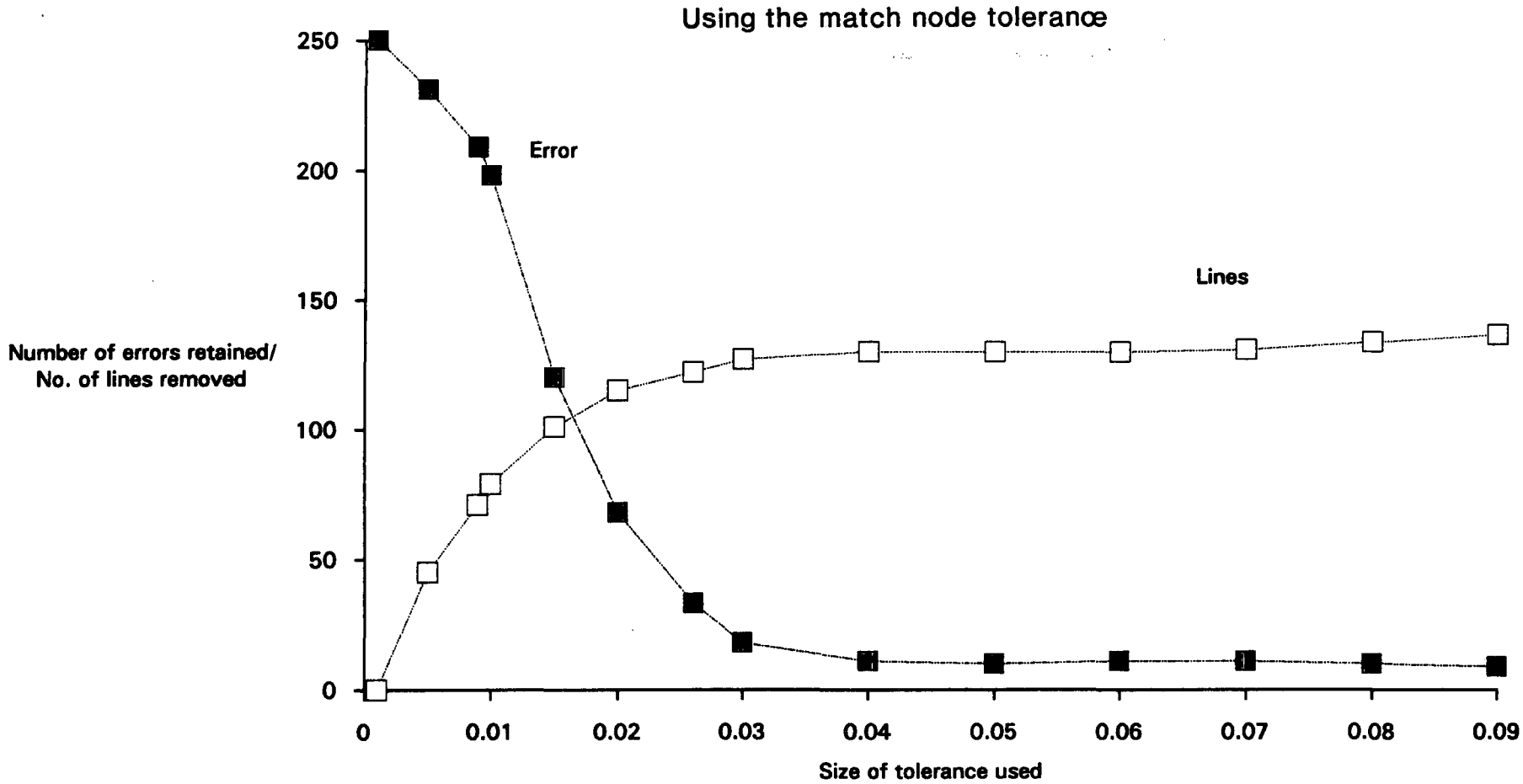


5.47 Subset of Fenland coverage used for subsequent experiments

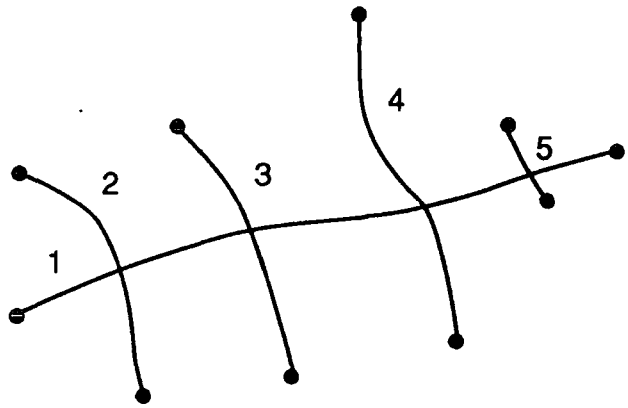
5.48 Number of errors removed by various node matching tolerances



5.49 Errors Vs lines

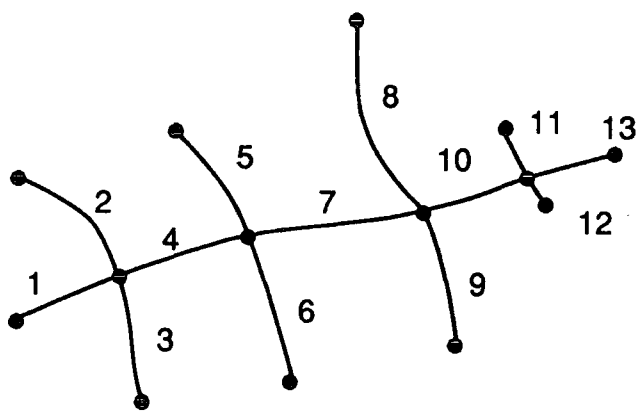


5.50 Problems of dangle removal following the use of the intersection program



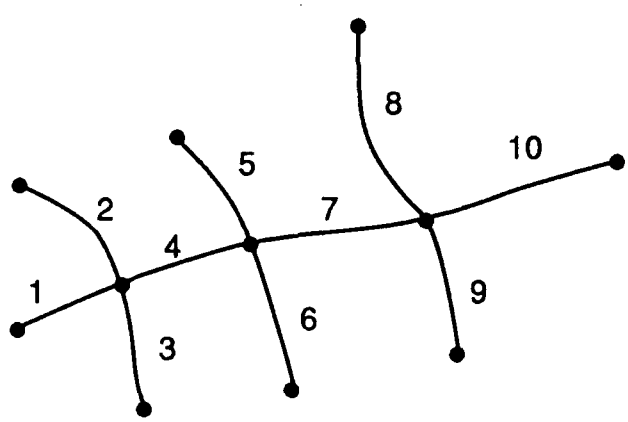
Pre-tolerancing.

Total number of lines: 5



Stage 1: Intersection

Total number of lines: 13

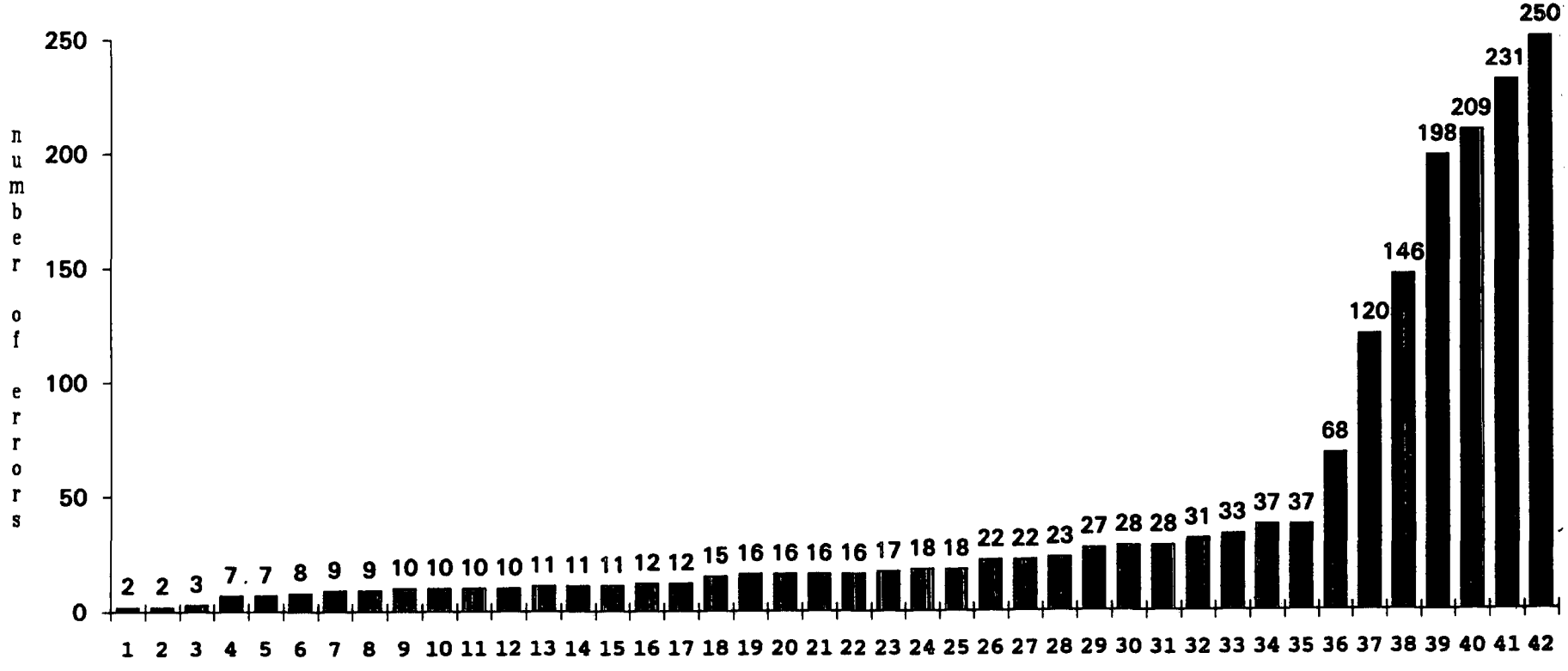


Stage 2: Removal of dangles

Total number of lines : 10

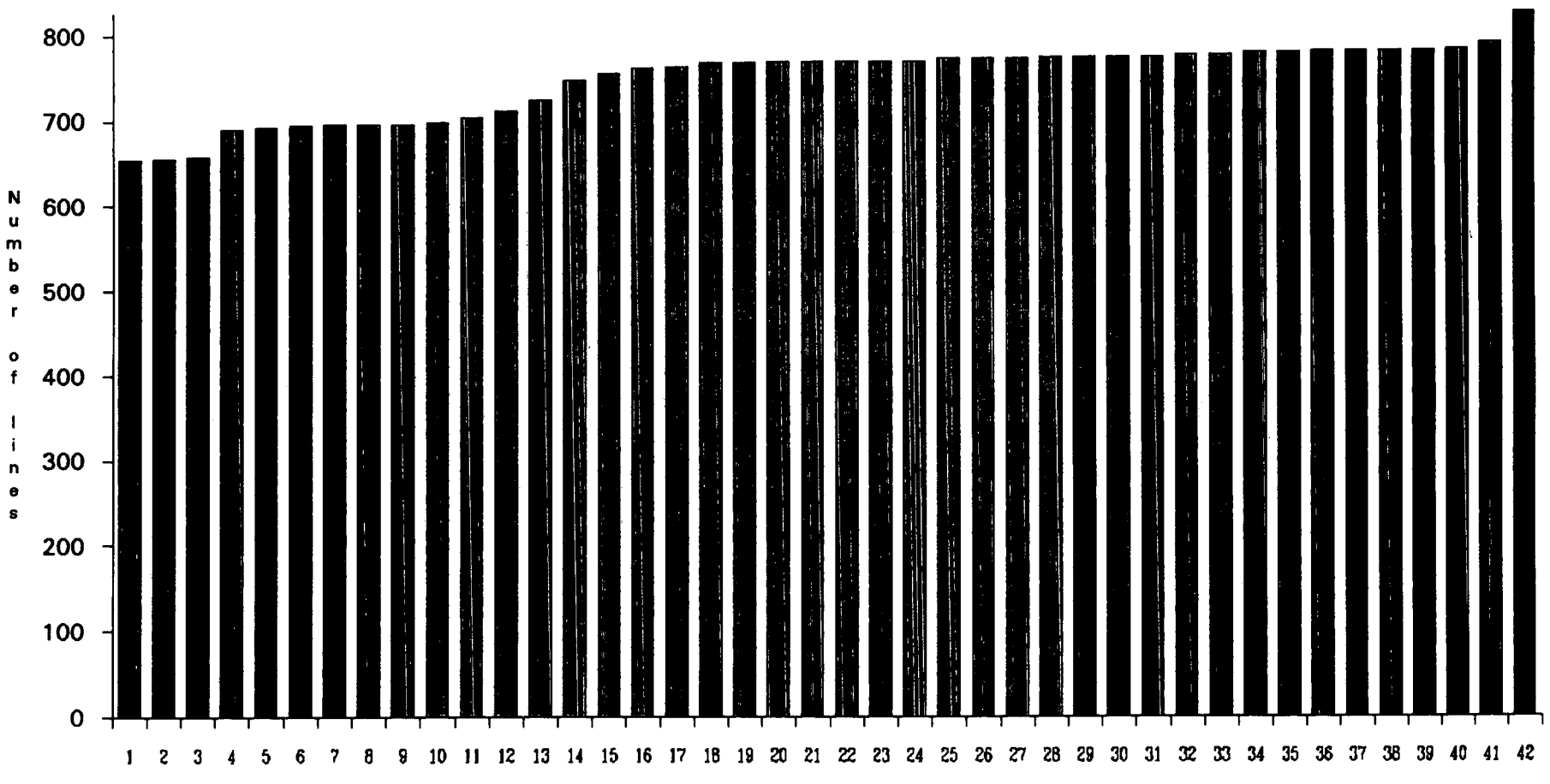
Removal of 'dangles' created by the intersection program cannot always be performed

5.51 The distribution of the total number of errors in the coverages following processing

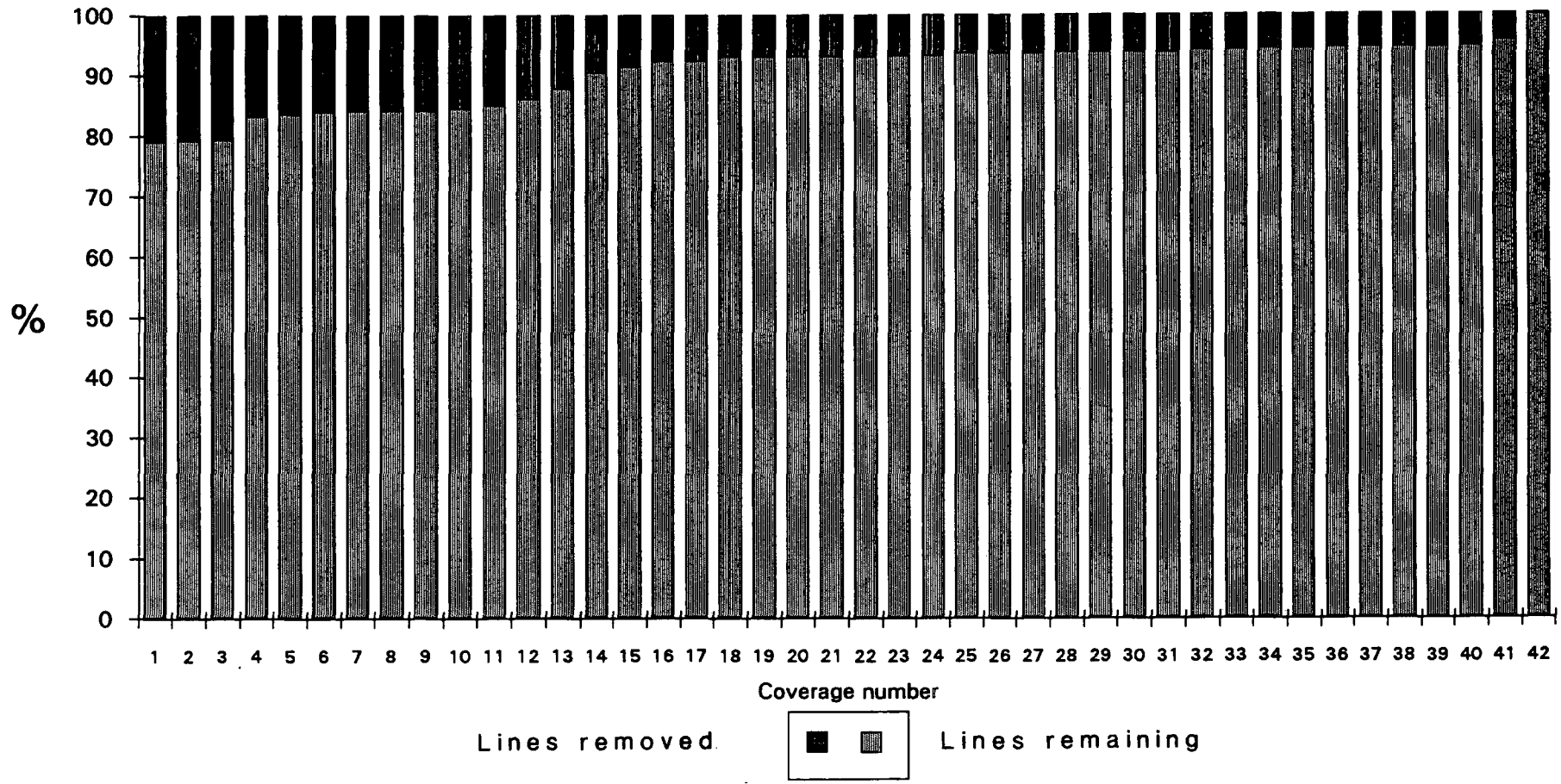


Cov number	Cov. n
1	T03C0
2	T09C2
3	T009C
4	T009C
5	T09C2
6	T04C2
7	T04C0
8	T09
9	T04C0
10	T04C0
11	T05
12	T08
13	T04
14	T06
15	T07
16	T03C2
17	T09C0
18	T03C0
19	T03C0
20	T03C0
21	T04C0
22	T04C0
23	T04C0
24	T03
25	T03C0
26	T0300
27	T03C0
28	T026C
29	T026C
30	T026C
31	T026C
32	T026C
33	T026
34	T026C
35	T026C
36	T02
37	T015
38	T009C
39	T01
40	T009
41	T005
42	T001

5.52 The distribution of the total number of lines in coverages following processing

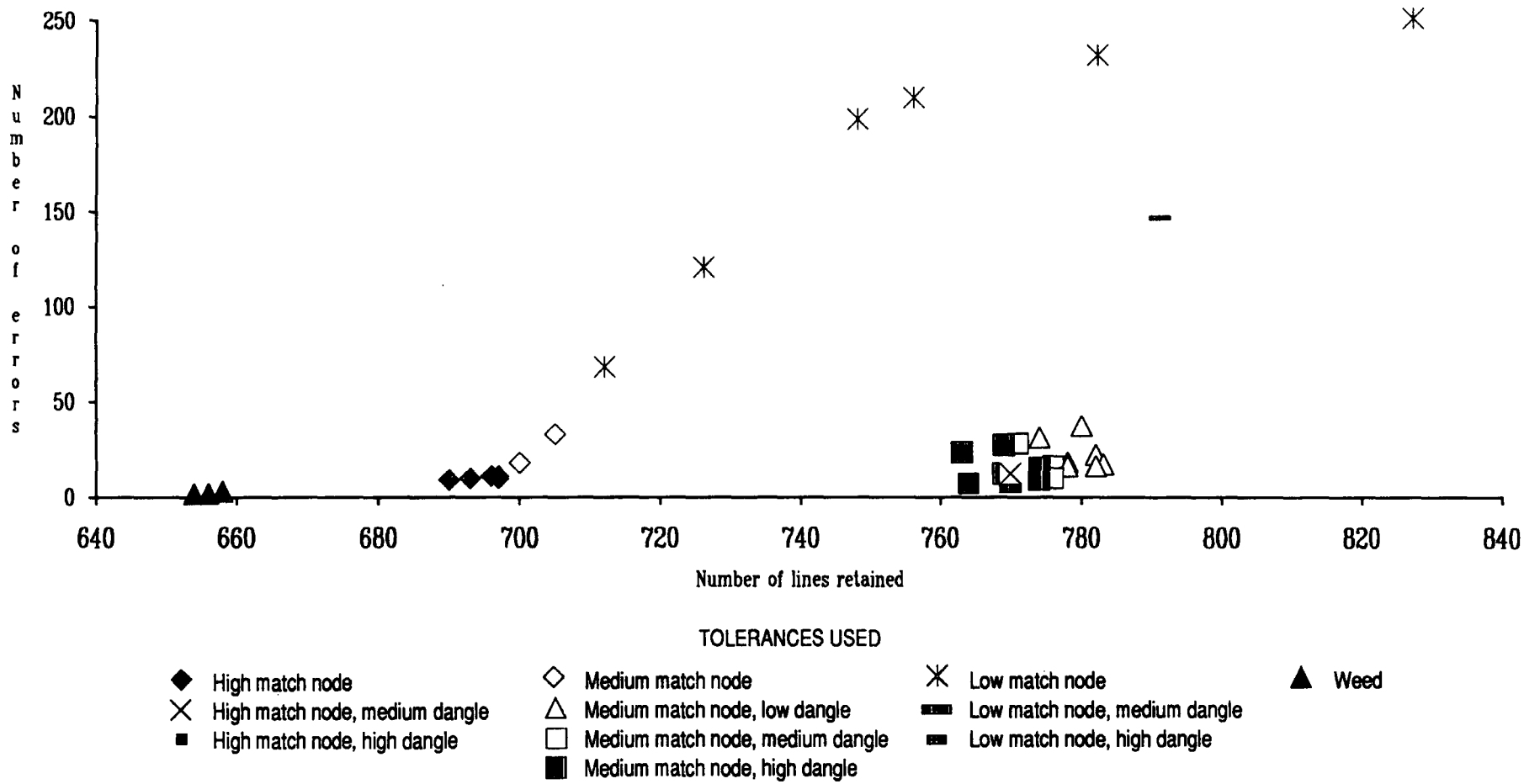


5.53 The percentage of lines remaining with respect to the lines removed

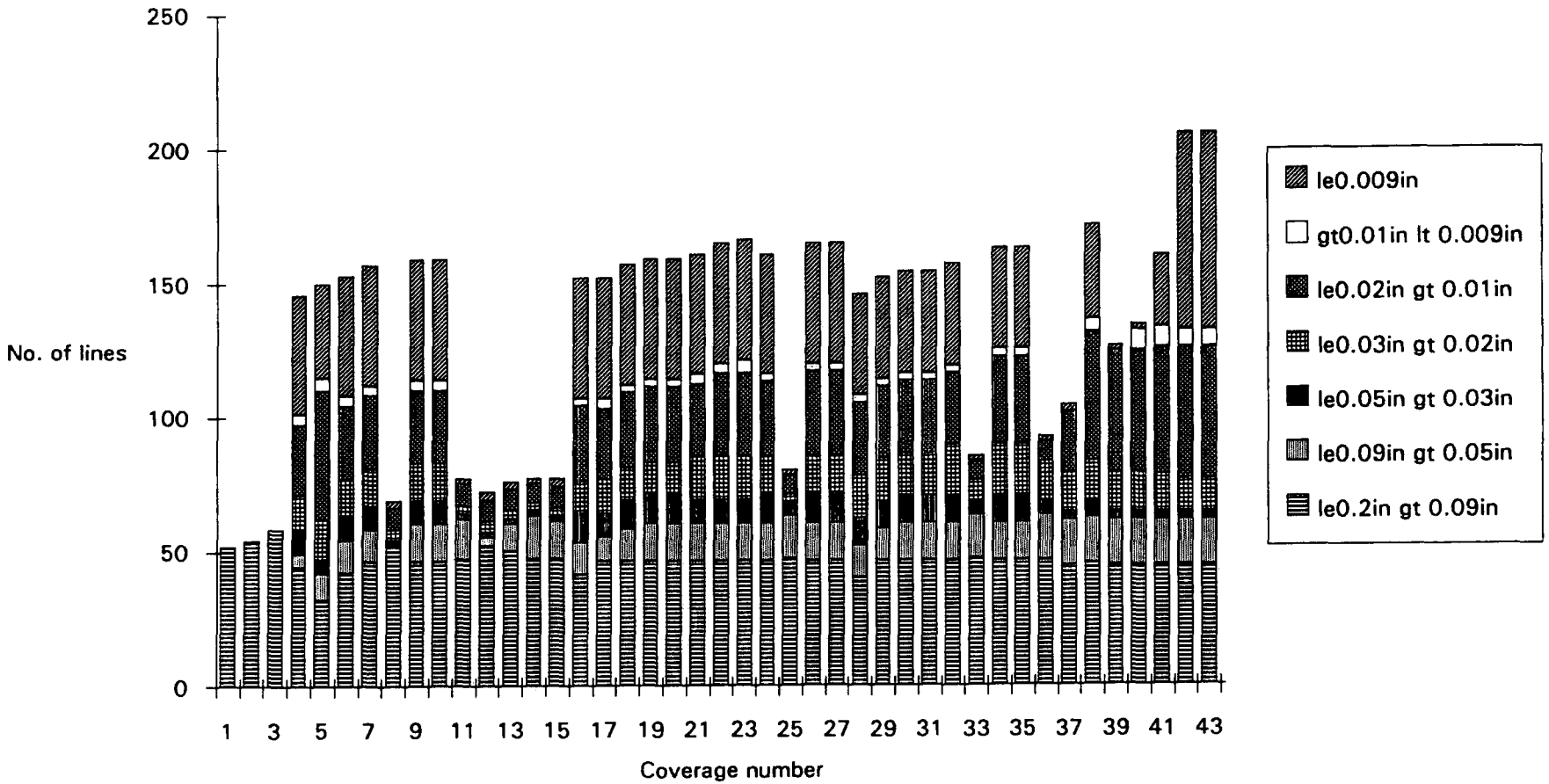


Cov.No	Cov. nar
1	T09C2V
2	T03C03V
3	T009C2V
4	T09
5	T08
6	T07
7	T05
8	T04
9	T06
10	T03
11	T026
12	T02
13	T015
14	T01
15	T009
16	T026C
17	T09C2
18	T03C2
19	T026C0
20	T009C
21	T04C2
22	T09C0
23	T026C0
24	T026C0
25	T04C0
26	T03C0
27	T026C0
28	T04C0
29	T04C0
30	T03C0
31	T03C0
32	T04C0
33	T03C0
34	T026C0
35	T026C0
36	T04C0
37	T0300
38	T03C0
39	T005
40	T04C00
41	T009C0
42	T001

5.54 A plot of the total number of errors retained as compared with the total number of lines retained

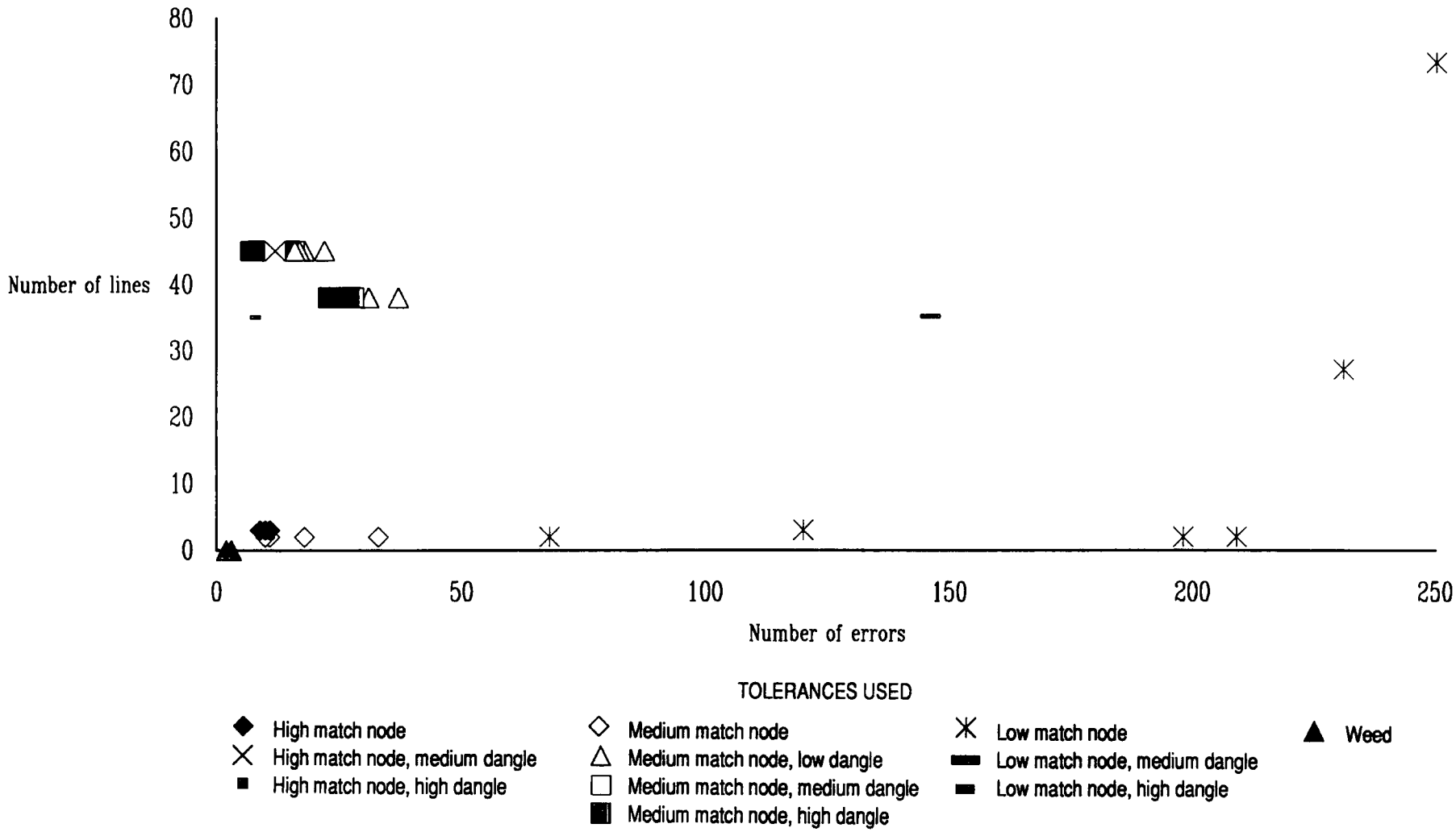


5.55 **Histogram of line distribution with respect to the type and size of tolerance used (sorted on the number of errors removed)**

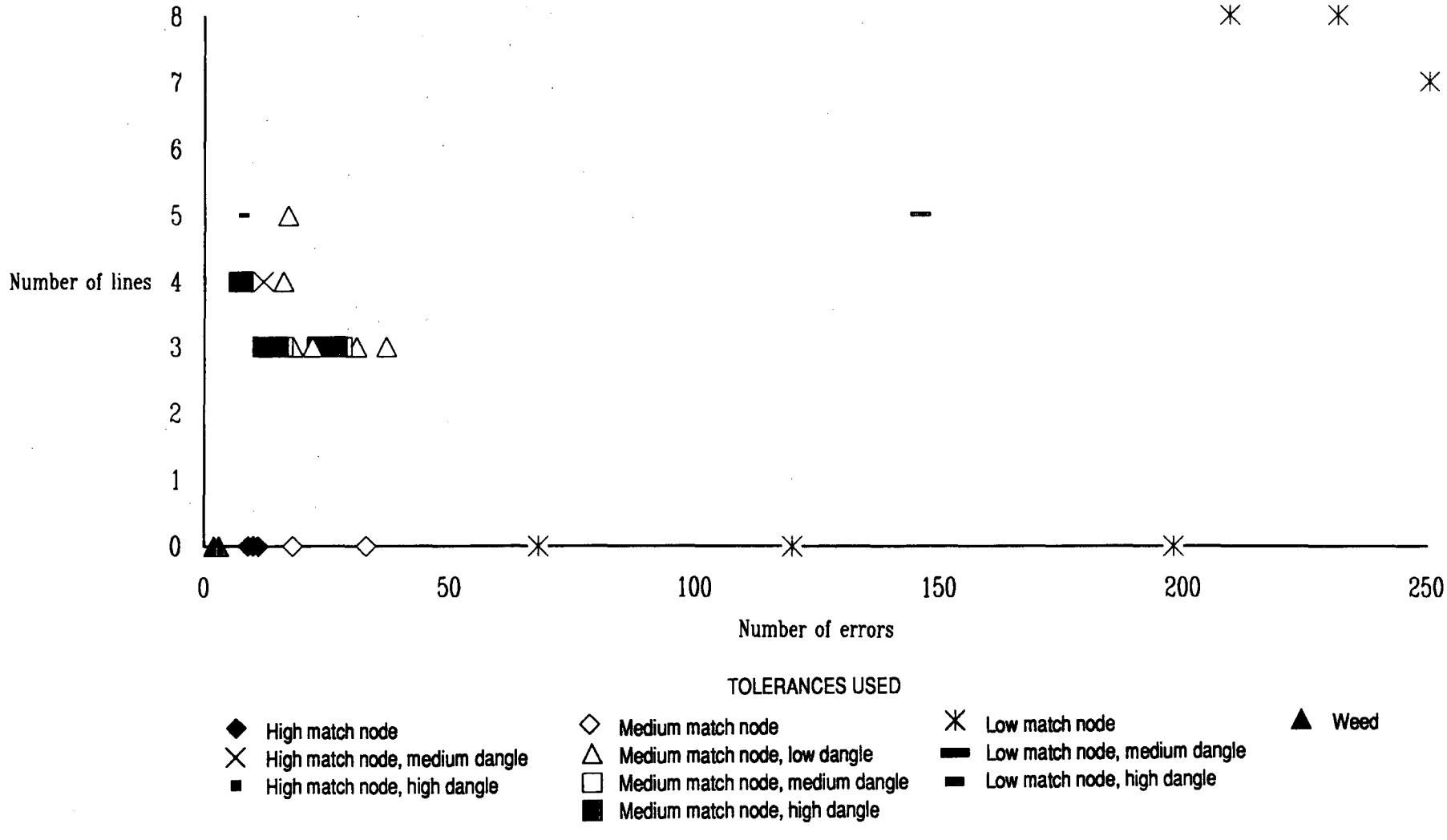


Cov. name	Cov.
T09C2W1	1
T03C03W1	2
T009C2W1	3
T09C2	4
T009C2	5
T04C2	6
T04C09	7
T09	8
T04C03	9
T04C05	1
T05	1
T08	1
T07	1
T04	1
T06	1
T03C2	1
T09C03	1
T03C09	1
T03C03	1
T03C05	2
T04C02	2
T04C01	2
T04C009	2
T03C02	2
T03	2
T03009	2
T03C01	2
T026C2	2
T026C09	2
T026C03	3
T026C05	3
T026C02	3
T026	3
T026C01	3
T026C009	3
T02	3
T015	3
T009C03	3
T01	3
T009	4
T005	4
T001	4
FENS1AAT	4

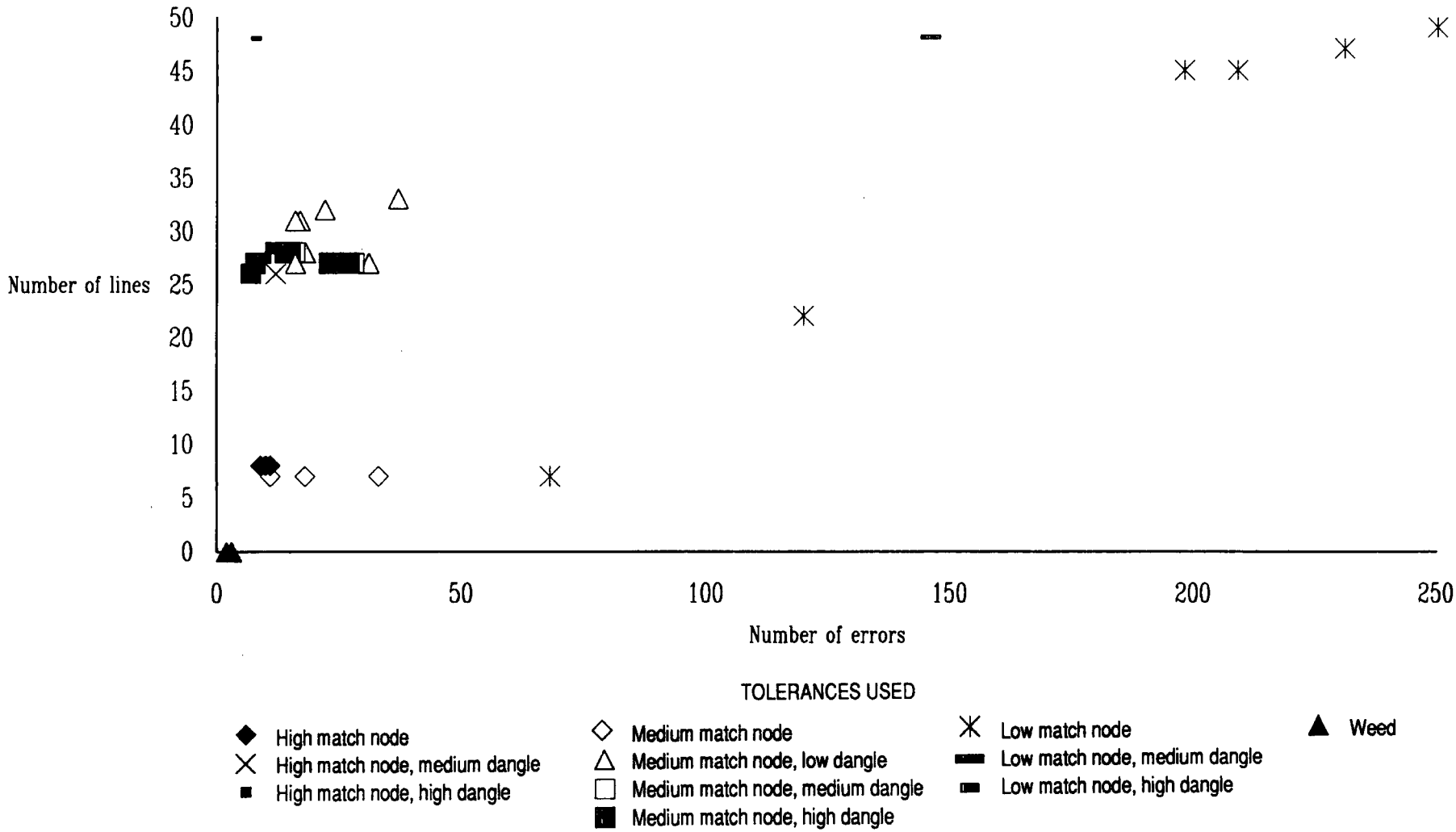
5.56 Lines less than 0.009in retained in the coverage



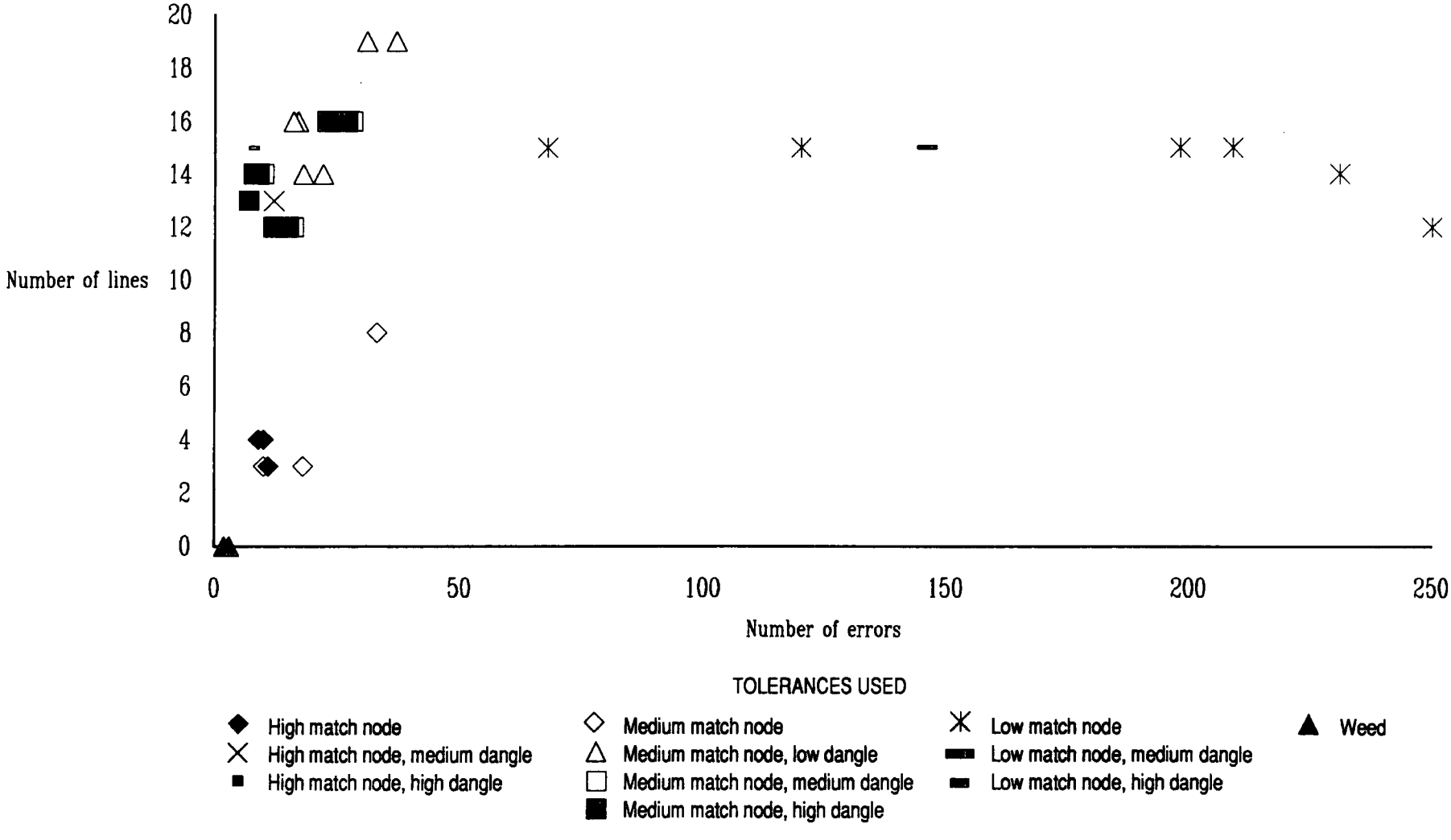
5.57 Lines greater than 0.009in and less than 0.01in retained in the coverage



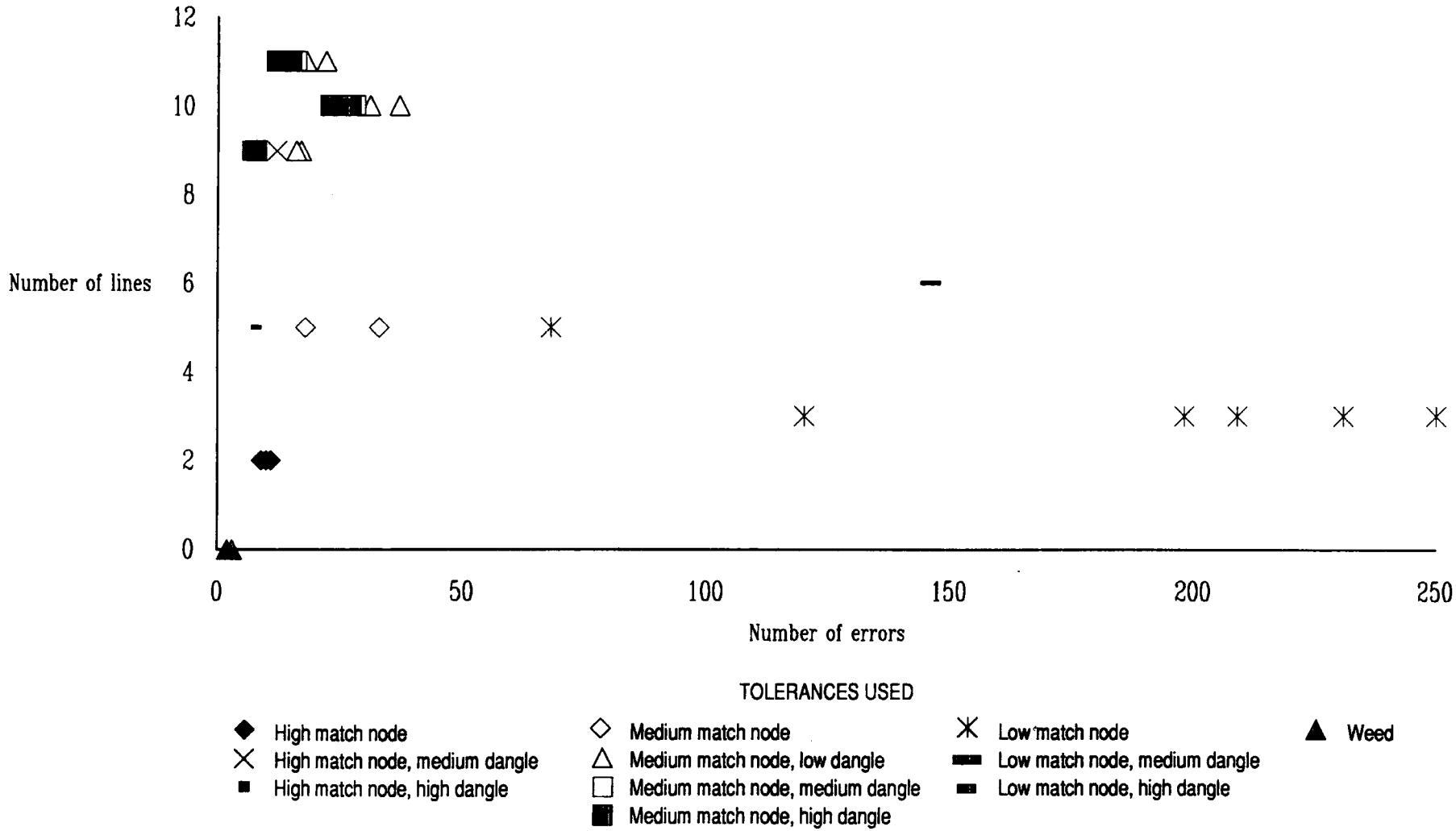
5.58 Lines greater than 0.01in and less than 0.02in retained in the coverage



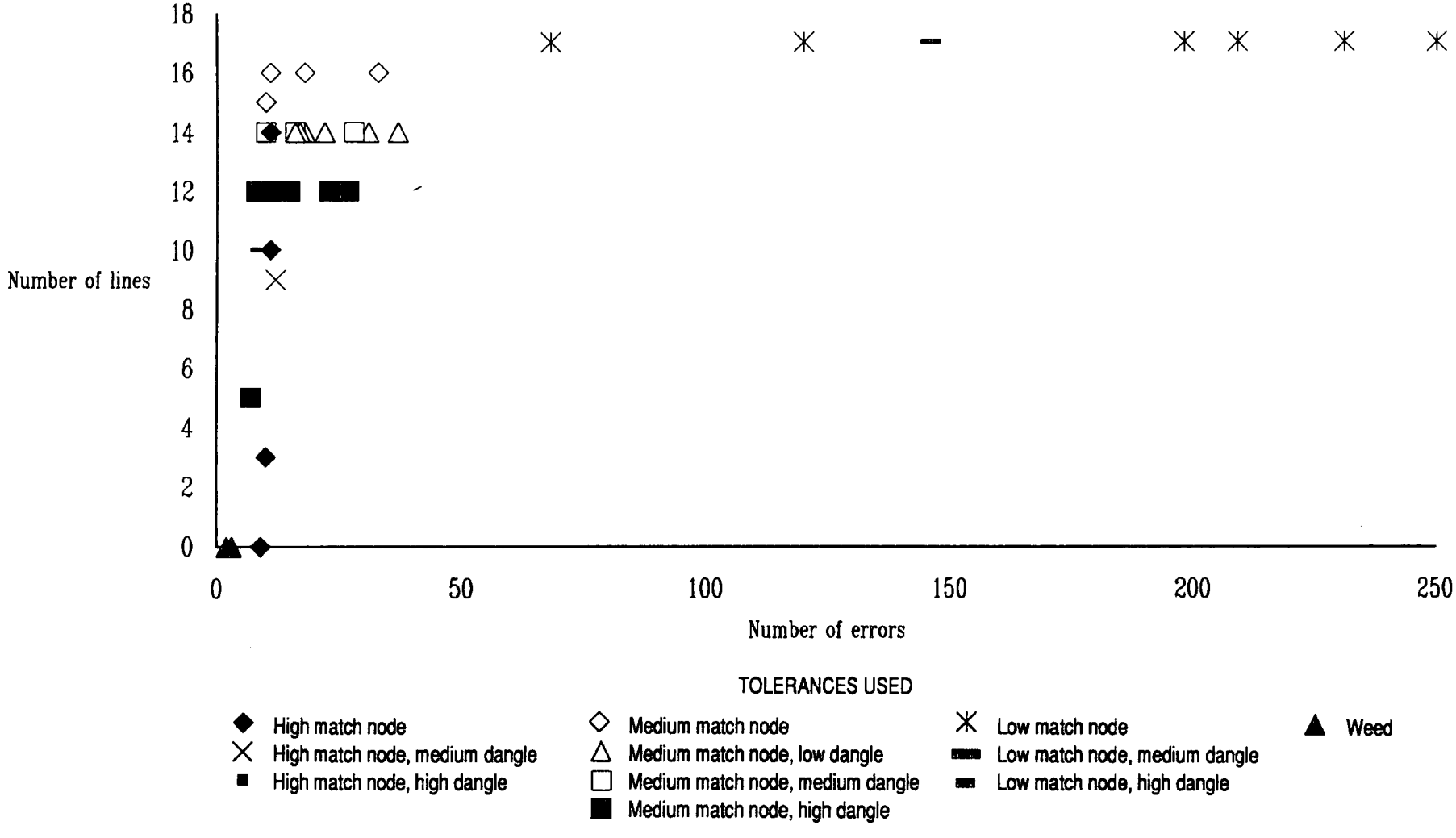
5.59 Lines greater than 0.02in and less than 0.03in retained in the coverage



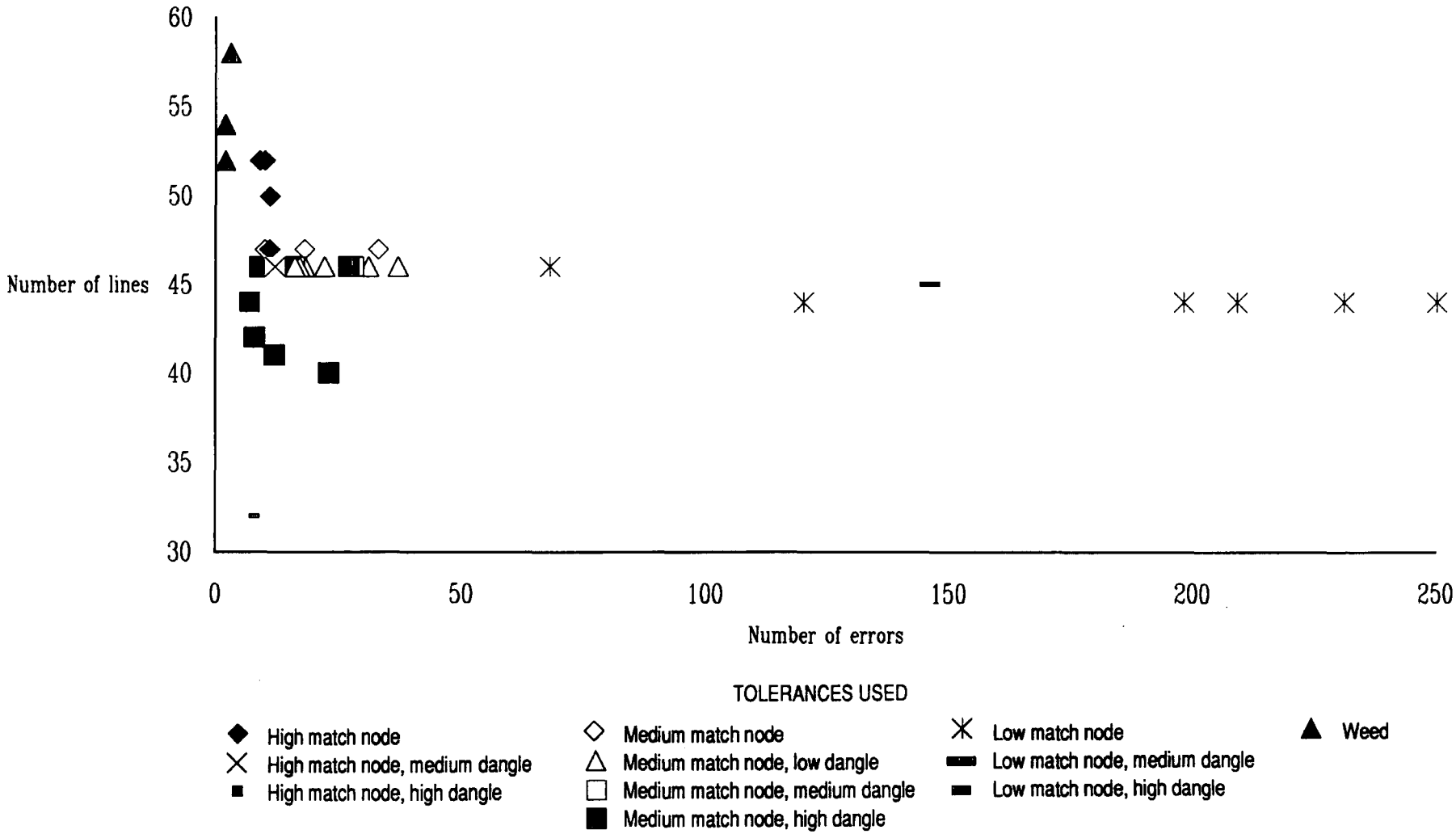
5.60 Lines greater than 0.03in and less than 0.05in retained in the coverage



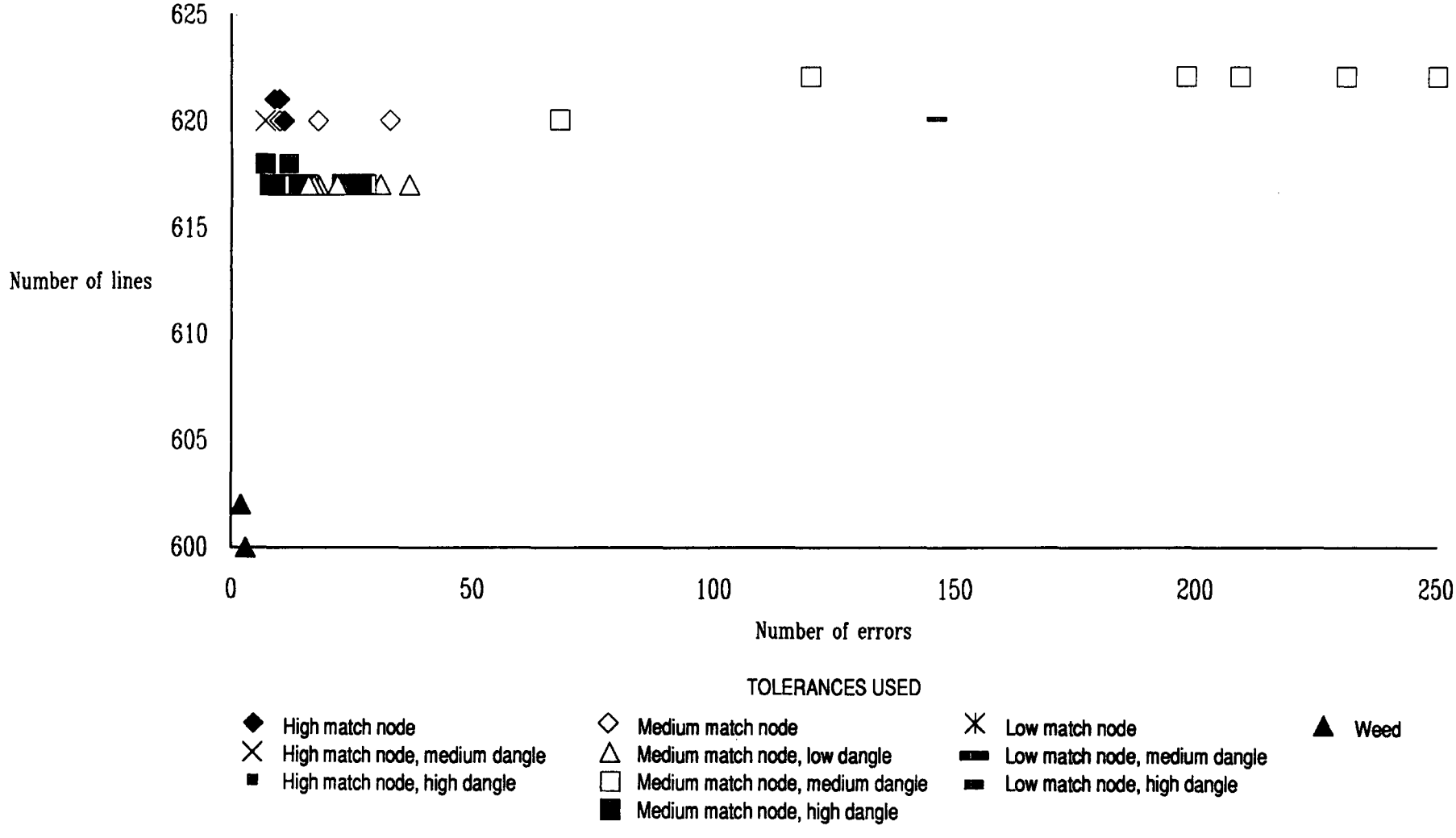
5.61 Lines greater than 0.05in and less than 0.09in retained in the coverage



5.62 Lines greater than 0.09in and less than 0.2in retained in the coverage



5.63 Lines greater than 0.2in retained in the coverage



5.64

Distribution of coverage line lengths following a match tol of 0.009" and a dangle tol of 0.03".

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  4.00 OCCURRENCES

    0        -.4
    0        -.2
  128         .0  **********************************
   98         .2  *************************
  115         .4  ******************************
  140         .6  ************************************
  104         .8  ***************************
   76        1.0  ********************
   55        1.2  **************
   30        1.4  ********
   17        1.6  ****
    3        1.8  *
    9        2.0  **
    5        2.2  *
    3        2.4  *
    2        2.6  *
    2        2.8  *
    3        3.0  *
    1        3.2
    0        3.4
    0        3.6
             I....+....I....+....I....+....I....+....I....+....I....+....I
             0        40        80       120       160       200
                          HISTOGRAM FREQUENCY
```

MEAN	.648	STD ERR	.019	MEDIAN	.577
MODE	.011	STD DEV	.528	VARIANCE	.279
KURTOSIS	2.810	S E KURT	.174	SKEWNESS	1.302
S E SKEW	.087	RANGE	3.219	MINIMUM	.002
MAXIMUM	3.221	SUM	512.318		

5.65

Distribution of coverage line lengths following a match tol of 0.009" and a dangle tol of 0.2"

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  4.00 OCCURRENCES

    0        -.4
    0        -.2
  120         .0  ******************************
   85         .2  *********************
  115         .4  ******************************
  140         .6  ************************************
  104         .8  ***************************
   76        1.0  ********************
   55        1.2  **************
   30        1.4  ********
   17        1.6  ****
    3        1.8  *
    9        2.0  **
    5        2.2  *
    3        2.4  *
    2        2.6  *
    2        2.8  *
    3        3.0  *
    1        3.2
    0        3.4
    0        3.6
             I....+....I....+....I....+....I....+....I....+....I....+....I
             0        40        80       120       160       200
                        HISTOGRAM FREQUENCY
```

MEAN	.662	STD ERR	.019	MEDIAN	.598
MODE	.011	STD DEV	.528	VARIANCE	.279
KURTOSIS	2.805	S E KURT	.176	SKEWNESS	1.285
S E SKEW	.088	RANGE	3.219	MINIMUM	.002
MAXIMUM	3.221	SUM	509.751		

5.66

Distribution of coverage line lengths following a match tol of 0.09" and a dangle tol of 0.03".

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  4.00 OCCURRENCES

    0        -.4
    0        -.2
  107         .0  ***************************
   99         .2  *************************
  111         .4  ****************************
  144         .6  ************************************
  100         .8  *************************
   77        1.0  ********************
   57        1.2  ***************
   28        1.4  *******
   18        1.6  *****
    4        1.8  *
    9        2.0  **
    5        2.2  *
    3        2.4  *
    2        2.6  *
    2        2.8  *
    3        3.0  *
    1        3.2
    0        3.4
    0        3.6
                I....+....I....+....I....+....I....+....I....+....I....+....I
                0        40        80       120       160       200
                          HISTOGRAM FREQUENCY
```

MEAN	.668	STD ERR	.019	MEDIAN	.604
MODE	.003	STD DEV	.527	VARIANCE	.278
KURTOSIS	2.786	S E KURT	.176	SKEWNESS	1.286
S E SKEW	.088	RANGE	3.219	MINIMUM	.002
MAXIMUM	3.221	SUM	514.176		

5.67

Distribution of coverage line lengths following a match tol of 0.026" and a dangle tol of 0.02".

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  4.00 OCCURRENCES

    0        -.4
    0        -.2
  113         .0  ****************************
   97         .2  ************************
  115         .4  ****************************
  139         .6  ***********************************
  103         .8  **************************
   75        1.0  ******************
   57        1.2  **************
   30        1.4  ********
   17        1.6  ****
    3        1.8  *
    9        2.0  **
    5        2.2  *
    3        2.4  *
    2        2.6  *
    2        2.8  *
    3        3.0  *
    1        3.2
    0        3.4
    0        3.6
                I....+....I....+....I....+....I....+....I....+....I....+....I
                0        40        80       120       160       200
                         HISTOGRAM FREQUENCY
```

MEAN	.662	STD ERR	.019	MEDIAN	.588
MODE	.003	STD DEV	.527	VARIANCE	.277
KURTOSIS	2.809	S E KURT	.176	SKEWNESS	1.297
S E SKEW	.088	RANGE	3.219	MINIMUM	.002
MAXIMUM	3.221	SUM	512.572		

5.68

Distribution of coverage line lengths following a match tol of 0.026" and a dangle tol of 0.2".

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  4.00 OCCURRENCES

    0        -.4
    0        -.2
  107         .0  ****************************
   92         .2  ************************
  115         .4  ******************************
  139         .6  ************************************
  103         .8  ***************************
   75        1.0  ********************
   57        1.2  ***************
   30        1.4  ********
   17        1.6  ****
    3        1.8  *
    9        2.0  **
    5        2.2  *
    3        2.4  *
    2        2.6  *
    2        2.8  *
    3        3.0  *
    1        3.2
    0        3.4
    0        3.6
                  I....+....I....+....I....+....I....+....I....+....I....+....I
                  0        40        80       120       160       200
                           HISTOGRAM FREQUENCY
```

MEAN	.670	STD ERR	.019	MEDIAN	.607
MODE	.003	STD DEV	.526	VARIANCE	.277
KURTOSIS	2.817	S E KURT	.177	SKEWNESS	1.292
S E SKEW	.089	RANGE	3.219	MINIMUM	.002
MAXIMUM	3.221	SUM	511.384		

5.69

Distribution of coverage line lengths following a match tol of 0.09" and a dangle tol of 0.2"

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  4.00 OCCURRENCES

    0        -.4
    0        -.2
  103         .0  *************************
   97         .2  ************************
  111         .4  ***************************
  144         .6  ************************************
  100         .8  *************************
   77        1.0  ******************
   57        1.2  **************
   28        1.4  *******
   18        1.6  *****
    4        1.8  *
    9        2.0  **
    5        2.2  *
    3        2.4  *
    2        2.6  *
    2        2.8  *
    3        3.0  *
    1        3.2
    0        3.4
    0        3.6
                  I....+....I....+....I....+....I....+....I....+....I....+....I
                  0        40        80       120       160       200
                           HISTOGRAM FREQUENCY
```

MEAN	.672	STD ERR	.019	MEDIAN	.609
MODE	.003	STD DEV	.527	VARIANCE	.277
KURTOSIS	2.791	S E KURT	.177	SKEWNESS	1.283
S E SKEW	.088	RANGE	3.219	MINIMUM	.002
MAXIMUM	3.221	SUM	513.525		

5.70

Distribution of coverage line lengths following a match tol of 0.009", a dangle tol of 0.2" and a weed tol of 0.1"

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  4.00 OCCURRENCES

   65        .15  ****************
   83        .30  ********************
   98        .45  ************************
  111        .60  ****************************
   82        .75  ********************
   69        .90  ******************
   46       1.05  ************
   37       1.20  *********
   32       1.35  ********
   12       1.50  ***
   11       1.65  ***
    2       1.80  *
    6       1.95  **
    1       2.10
    0       2.25
    1       2.40
    0       2.55
    0       2.70
    0       2.85
    1       3.00
    1       3.15
                I....+....I....+....I....+....I....+....I....+....I....+....I
                0        40        80       120       160       200
                          HISTOGRAM FREQUENCY
```

MEAN	.708	STD ERR	.017	MEDIAN	.640
MODE	.698	STD DEV	.424	VARIANCE	.180
KURTOSIS	2.996	S E KURT	.190	SKEWNESS	1.211
S E SKEW	.095	RANGE	3.122	MINIMUM	.101
MAXIMUM	3.223	SUM	465.745		

Distribution of coverage line lengths following a match tol of 0.09",
a dangle tol of 0.2" and a weed tol of 0.1"

COUNT	MIDPOINT	ONE SYMBOL EQUALS APPROXIMATELY 4.00 OCCURRENCES
61	.15	***************
88	.30	**********************
95	.45	************************
111	.60	****************************
81	.75	********************
68	.90	*****************
41	1.05	**********
40	1.20	**********
34	1.35	*********
11	1.50	***
12	1.65	***
2	1.80	*
6	1.95	**
1	2.10	
0	2.25	
1	2.40	
0	2.55	
0	2.70	
0	2.85	
1	3.00	
1	3.15	

I....+....I....+....I....+....I....+....I....+....I....+....I
0 40 80 120 160 200
HISTOGRAM FREQUENCY

MEAN	.711	STD ERR	.017	MEDIAN	.642
MODE	.385	STD DEV	.424	VARIANCE	.180
KURTOSIS	2.980	S E KURT	.191	SKEWNESS	1.221
S E SKEW	.096	RANGE	3.117	MINIMUM	.104
MAXIMUM	3.221	SUM	465.181		

5.72

A frequency distribution of the lengths in the T026c2 coverage following a square root transformation on the data.

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  4.00 OCCURRENCES

    0        -.1
    5         .0  *
   66         .1  *****************
   25         .2  ******
   18         .3  *****
   33         .4  ********
   53         .5  *************
   61         .6  ***************
  103         .7  ***************************
  102         .8  ***************************
   91         .9  ************************
   75        1.0  ******************
   62        1.1  ****************
   29        1.2  *******
   14        1.3  ****
   10        1.4  ***
    6        1.5  **
    4        1.6  *
    5        1.7  *
    1        1.8
    0        1.9
                 I....+....I....+....I....+....I....+....I....+....I....+....I
                 0        40        80       120       160       200
                          HISTOGRAM FREQUENCY
```

MEAN	.741	STD ERR	.013	MEDIAN	.779
MODE	.055	STD DEV	.348	VARIANCE	.121
KURTOSIS	-.117	S E KURT	.177	SKEWNESS	-.136
S E SKEW	.089	RANGE	1.750	MINIMUM	.045
MAXIMUM	1.795	SUM	565.425		

5.73

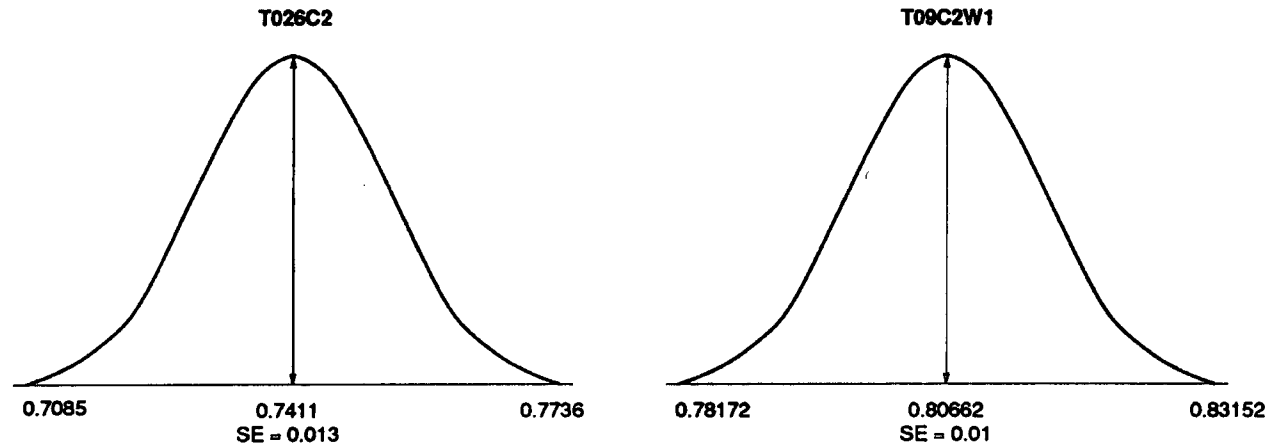
A frequency distribution of the lengths in the T09c2w1 coverage following a square root transformation on the data.

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  2.00 OCCURRENCES

    4       .300  **
   31       .375  ****************
   32       .450  ****************
   52       .525  ***************************
   47       .600  ************************
   66       .675  **********************************
   77       .750  ****************************************
   88       .825  *********************************************
   69       .900  ***********************************
   54       .975  ****************************
   41      1.050  *********************
   42      1.125  *********************
   24      1.200  ************
   15      1.275  ********
    3      1.350  **
    6      1.425  ***
    1      1.500  *
    0      1.575
    0      1.650
    1      1.725  *
    1      1.800  *
                 I....+....I....+....I....+....I....+....I....+....I
                 0        20        40        60        80       100
                          HISTOGRAM FREQUENCY
```

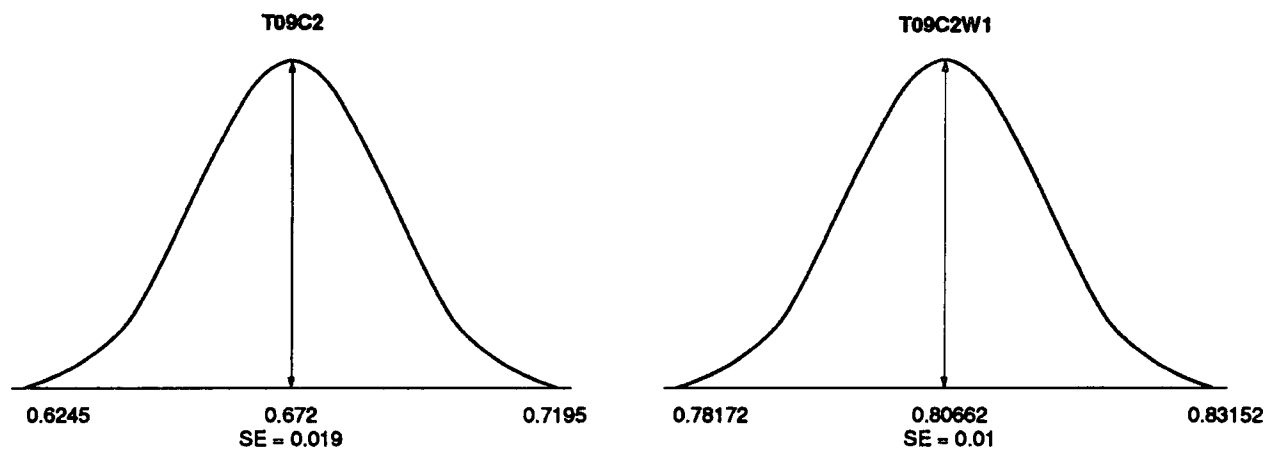
MEAN	.807	STD ERR	.010	MEDIAN	.801
MODE	.620	STD DEV	.246	VARIANCE	.061
KURTOSIS	-.001	S E KURT	.191	SKEWNESS	.298
S E SKEW	.096	RANGE	1.472	MINIMUM	.322
MAXIMUM	1.795	SUM	527.530		
VALID CASES	654	MISSING CASES	0		

5.74 A COMPARISON OF THE T-TEST RESULTS

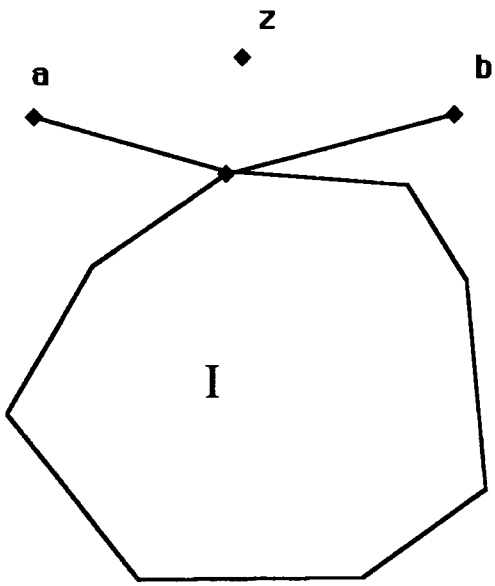
(a) COVERAGES T026C2 AND T09C2W1



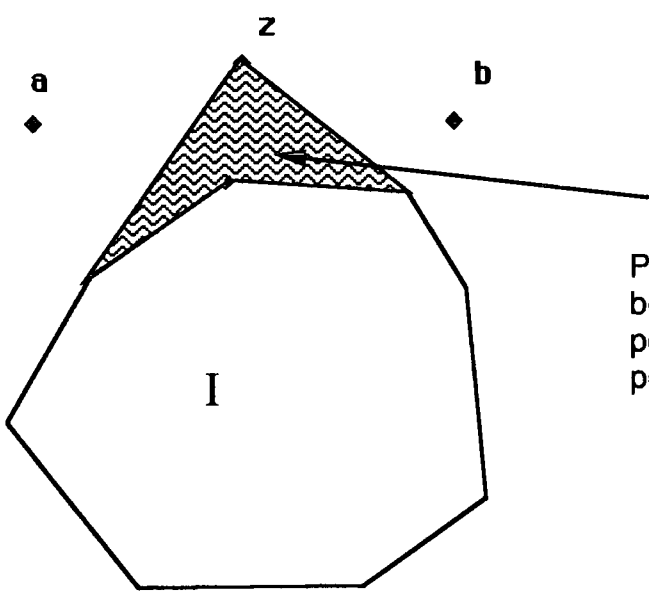
(b) COVERAGES T09C2 AND T09C2W1



5.75 Reduction of polygon size due to poor digitising

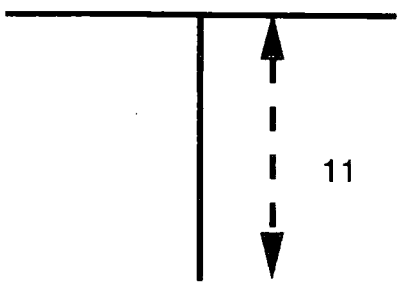


The intended stopping point of the line ends that make up polygon I is at point Z, however, since this point has being mistakenly identified on two occasions, the the size of the polygon has been reduced.

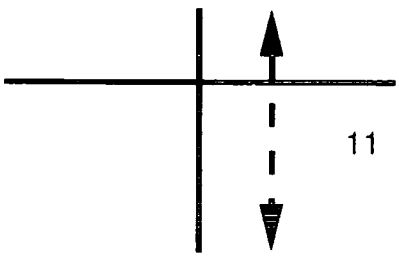


Potential areal difference between the intended polygon and the actual polygon

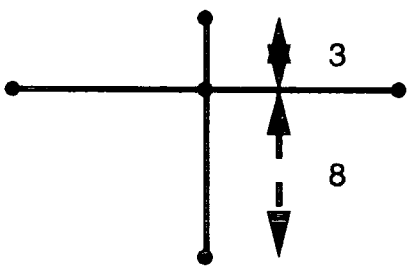
5.76 The destructive power of faulty intersecting



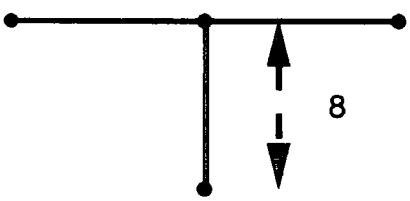
Original T-junction as it is viewed in the real world, with the stem carrying a length of 11 units.



The original is digitised and the stem is given the correct length of 11 units. However the horizontal intersection is mis-placed, thus the intended length will be shortened when the data model is built.

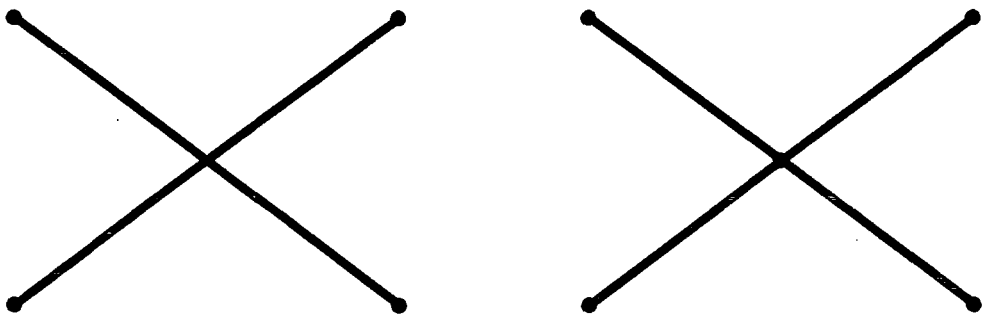


Stage one of building the data model. All lines that cross eachother are intersected, thus the t-junction now becomes a cross-road. As the line is intersected, its length is also narrowed, thus the length of what will be recognised as the stem has diminished to 8 units, and a redundant line of length 3 units remains.

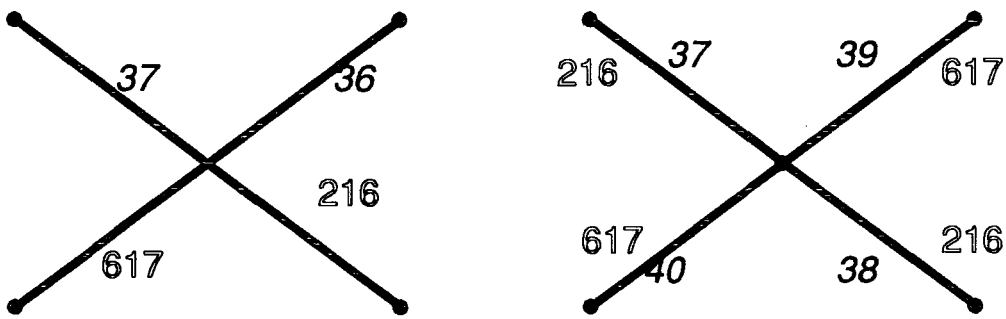


Stage two of building the data model. A tolerance is set to remove all 'dangling' arcs. In this case the tolerance is 3.5 units. The small line created in stage one of building the data model falls below this and is therefore removed. The result is that the stem is 3 units shorter that is should be, despite accuracy in its original digitised length.

5.77 The use of an additional identifier to track lines in ARC/INFO



Line number increase as part of the clean and build process in ARC/INFO

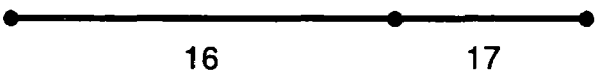


The use of an extra identifer was employed so that lines could be 'tracked'

Although the identifier of the line may change, the additional identifier remains constant, and follows the line as it is intersected, split and re-numbered

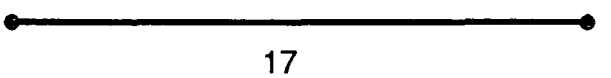
5.78 Evaluating the changes in line length following tolerancing

Original coverage



id	length
16	0.674in
17	0.004in

Test coverage

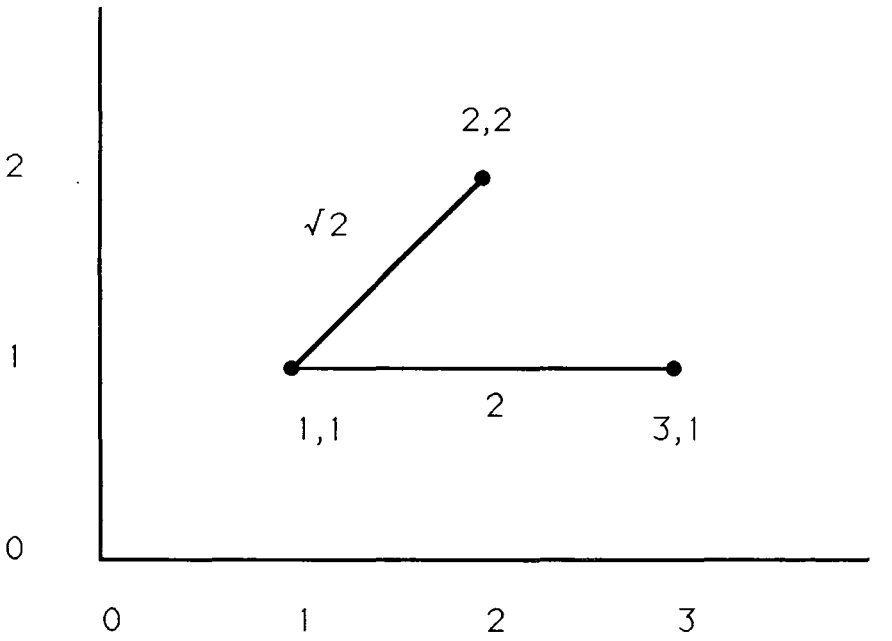


id	length
16	0in
17	0.678in

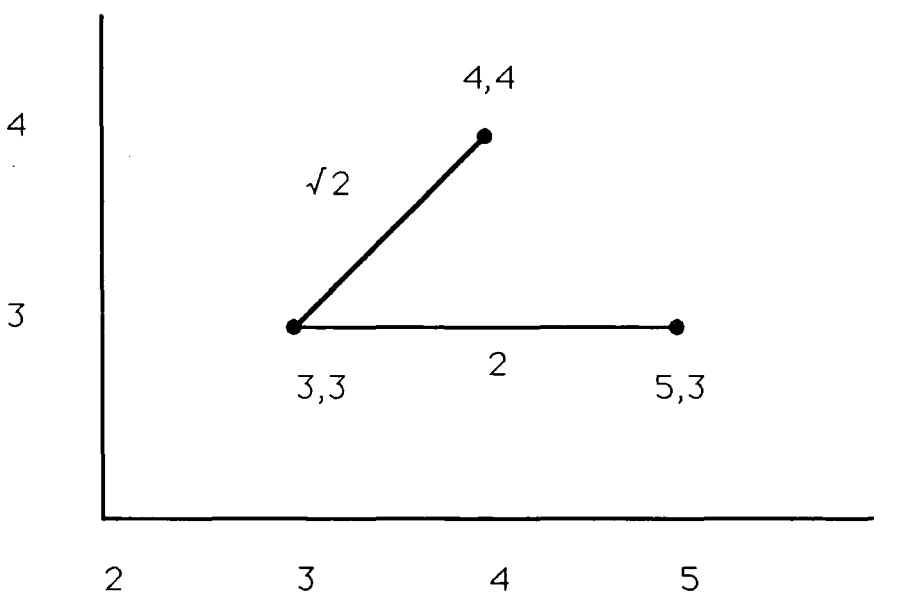
Test		Original	Difference
0in	-	0.674in	-0.674in
0.678in	-	0.004in	+0.674in
Sum:-			0 inches

Therefore, despite relabelling changes in the database, no positional change has occurred

5.79 Problems of line lengths and geometric movements



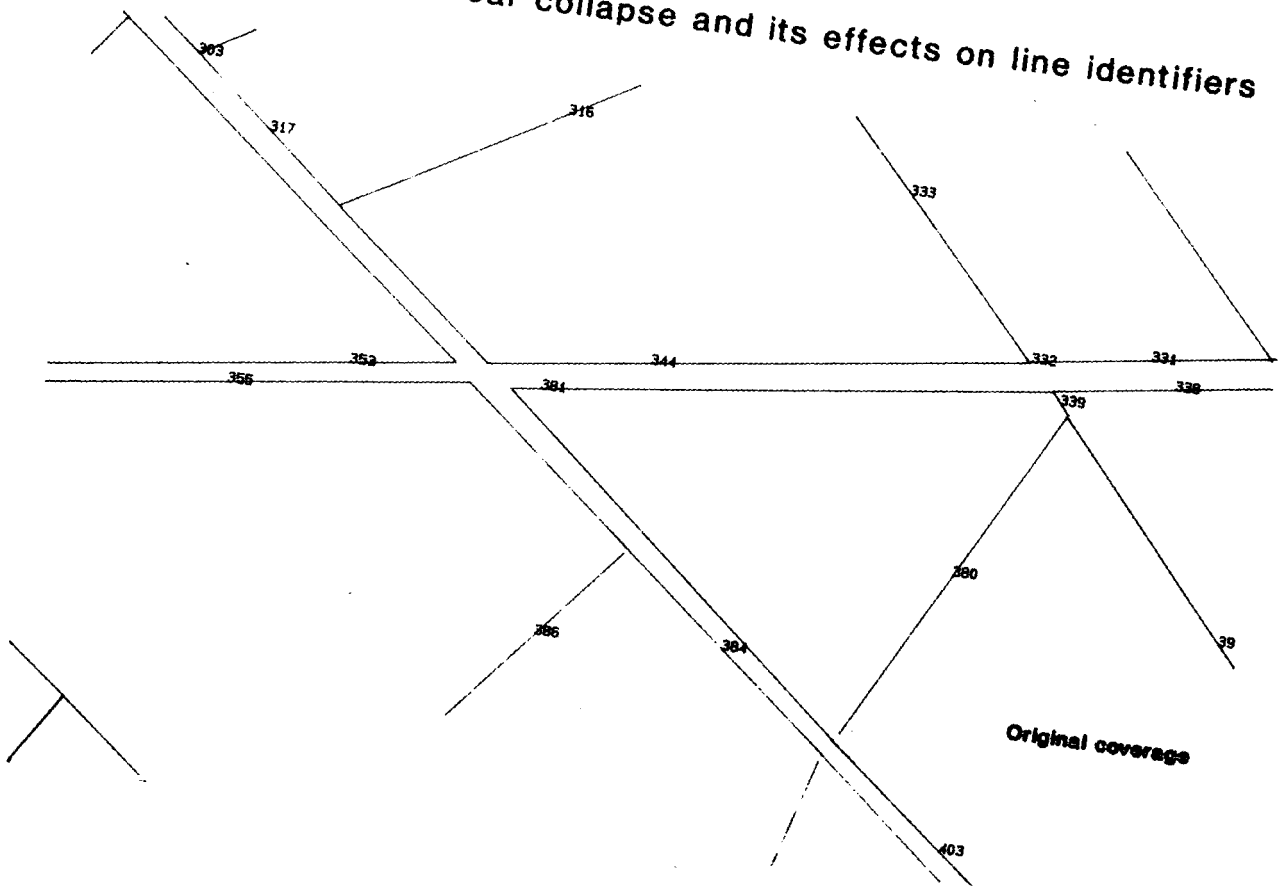
Original coverage



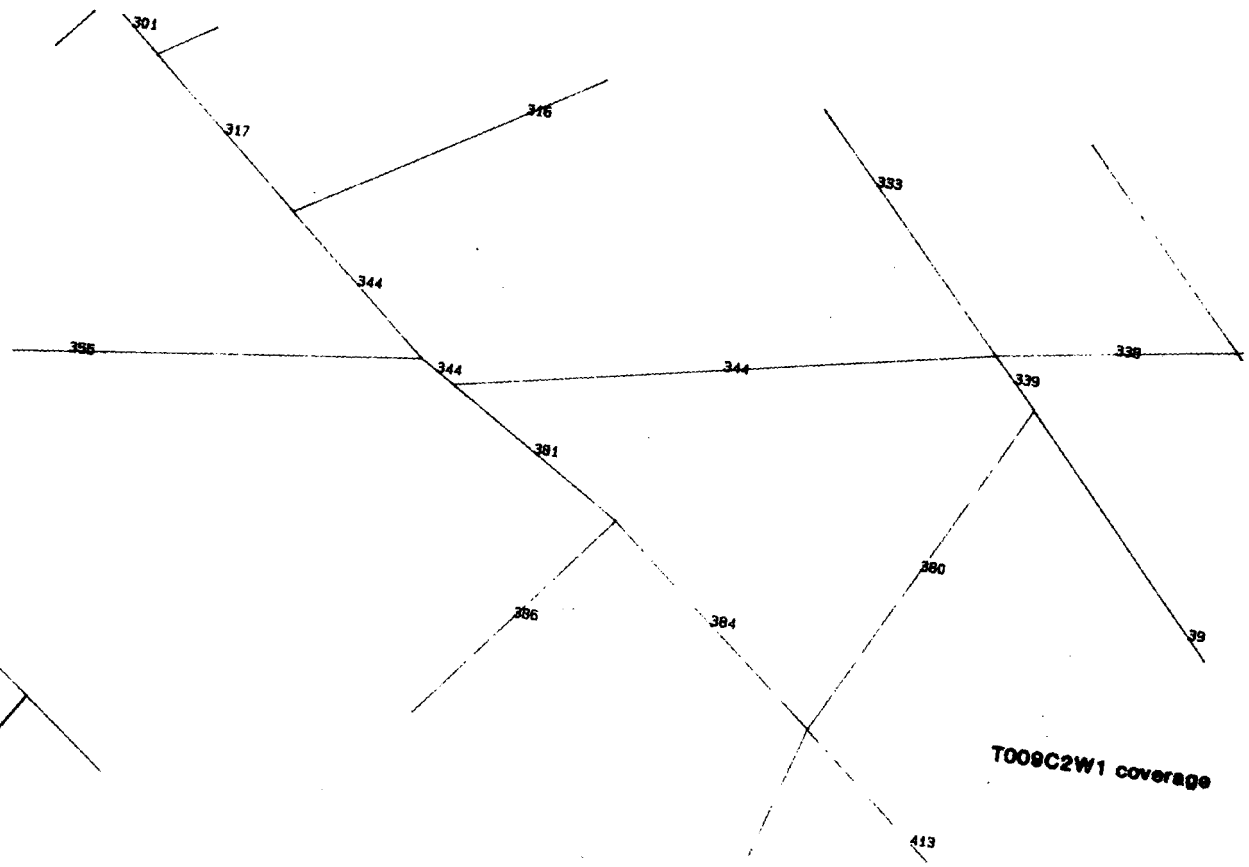
Coverage following tolerancing

Although the line lengths of the lines in question have remained the same, their relative position in space has changed. Therefore geometric change has taken place.

5.80 Linear collapse and its effects on line identifiers



Original coverage



T009C2W1 coverage

and T09C2

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  8.00 OCCURRENCES

  300     .000  ************************************
   99     .002  ************
   38     .004  *****
   14     .006  **
   19     .008  **
   20     .010  ***
   37     .012  *****
   37     .014  *****
   32     .016  ****
   30     .018  ****
   15     .020  **
   14     .022  **
    7     .024  *
   11     .026  *
    1     .028
    4     .030  *
    1     .032
    5     .034  *
    1     .036
    0     .038
    1     .040
          I....+....I....+....I....+....I....+....I....+....I....+....I
          0        80       160       240       320       400
                   HISTOGRAM FREQUENCY
```

TO26CO2

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  8.00 OCCURRENCES
  298     .000  ************************************
   97     .002  ************
   36     .004  *****
   14     .006  **
   19     .008  **
   19     .010  **
   36     .012  *****
   37     .014  *****
   32     .016  ****
   30     .018  ****
   13     .020  **
   13     .022  **
    7     .024  *
   12     .026  **
    1     .028
    5     .030  *
    1     .032
    5     .034  *
    1     .036
    0     .038
    1     .040
          I....+....I....+....I....+....I....+....I....+....I....+....I
          0        80       160       240       320       400
                   HISTOGRAM FREQUENCY
```

TO26C2

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY 10.00 OCCURRENCES

    0     -.04
    0     -.02
  402      .00  ****************************************
  186      .02  ******************
   26      .04  ***
   17      .06  **
   26      .08  ***
    2      .10
    0      .12
    1      .14
    0      .16
    0      .18
    0      .20
    0      .22
    1      .24
    0      .26
    0      .28
    0      .30
    1      .32
    0      .34
    0      .36
          I....+....I....+....I....+....I....+....I....+....I....+....I
          0       100       200       300       400       500
                   HISTOGRAM FREQUENCY
```

TO9C2

```
COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY 10.00 OCCURRENCES

    0     -.04
    0     -.02
  405      .00  ****************************************
  188      .02  ******************
   27      .04  ***
   16      .06  **
   26      .08  ***
    2      .10
    0      .12
    1      .14
    1      .16
    0      .18
    0      .20
    0      .22
    0      .24
    0      .26
    0      .28
    0      .30
    1      .32
    0      .34
    0      .36
          I....+....I....+....I....+....I....+....I....+....I....+....I
          0       100       200       300       400       500
                   HISTOGRAM FREQUENCY
```

T09C03

TO9C2W1 TO09C03, T009C2 and T009C2W1

COUNT MIDPOINT ONE SYMBOL EQUALS APPROXIMATELY 12.00 OCCURRENCES

0	-.40	
0	-.25	
0	-.10	
560	.05	**********
29	.20	**
7	.35	*
3	.50	
4	.65	
2	.80	
2	.95	
0	1.10	
0	1.25	
1	1.40	
1	1.55	
1	1.70	
0	1.85	
0	2.00	
1	2.15	
0	2.30	
0	2.45	
0	2.60	

T09C2W1

I....+....I....+....I....+....I....+....I....+....I....+....I
0 120 240 360 480 600
HISTOGRAM FREQUENCY

COUNT MIDPOINT ONE SYMBOL EQUALS APPROXIMATELY 12.00 OCCURRENCES

520	-.001	**********
139	.004	**********
33	.009	***
8	.014	*
7	.019	*
4	.024	
2	.029	
1	.034	
0	.039	
0	.044	
0	.049	
0	.054	
0	.059	
0	.064	
0	.069	
0	.074	
0	.079	
0	.084	
0	.089	
0	.094	
1	.099	

TOO9CO3

I....+....I....+....I....+....I....+....I....+....I....+....I
0 120 240 360 480 600
HISTOGRAM FREQUENCY

COUNT MIDPOINT ONE SYMBOL EQUALS APPROXIMATELY 12.00 OCCURRENCES

501	-.001	**********
138	.004	**********
32	.009	***
8	.014	*
7	.019	*
4	.024	
2	.029	
2	.034	
0	.039	
0	.044	
0	.049	
0	.054	
0	.059	
0	.064	
0	.069	
0	.074	
0	.079	
0	.084	
0	.089	
0	.094	
1	.099	

TOO9C2

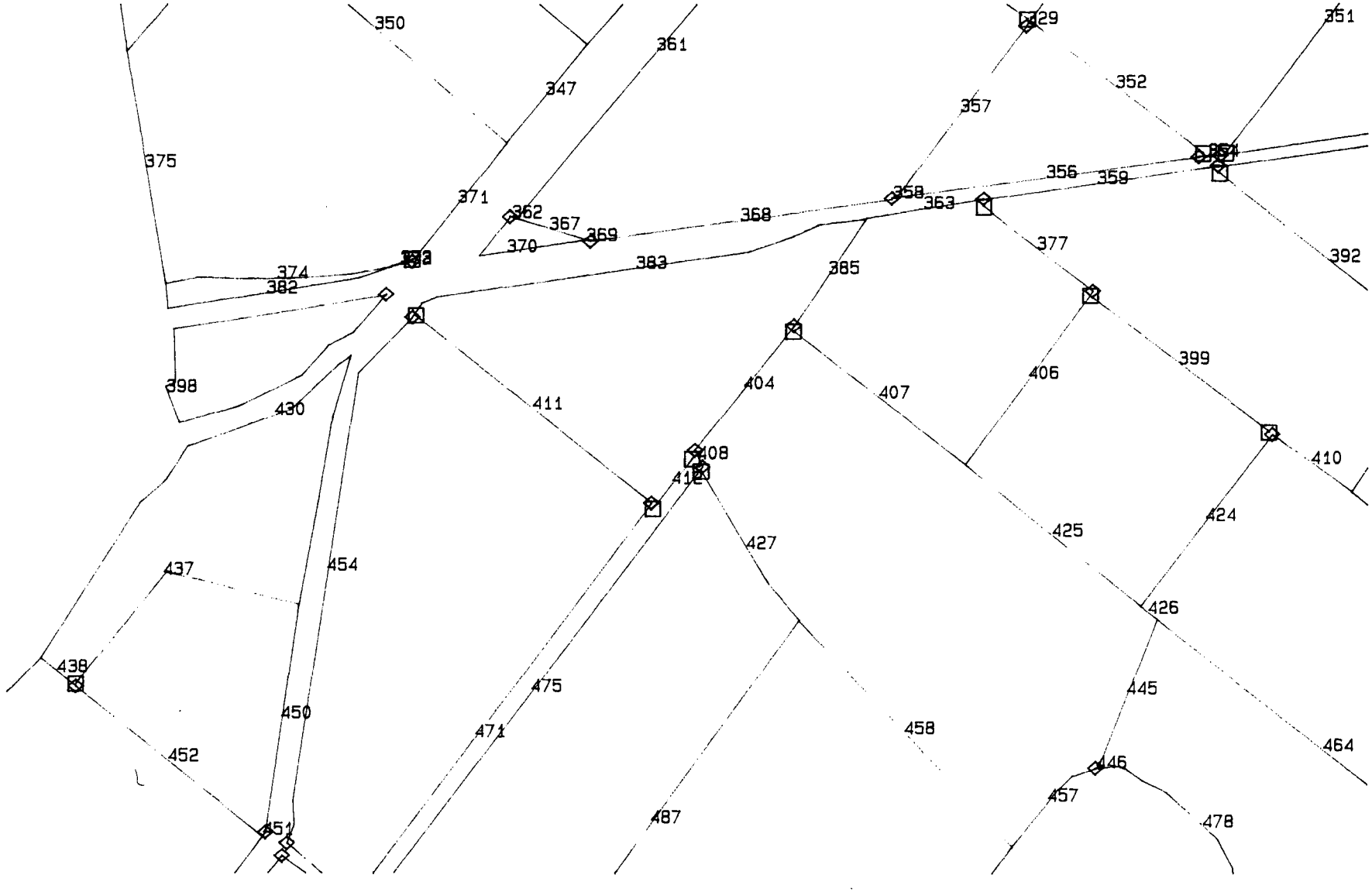
I....+....I....+....I....+....I....+....I....+....I....+....I
0 120 240 360 480 600
HISTOGRAM FREQUENCY

COUNT MIDPOINT ONE SYMBOL EQUALS APPROXIMATELY 12.00 OCCURRENCES

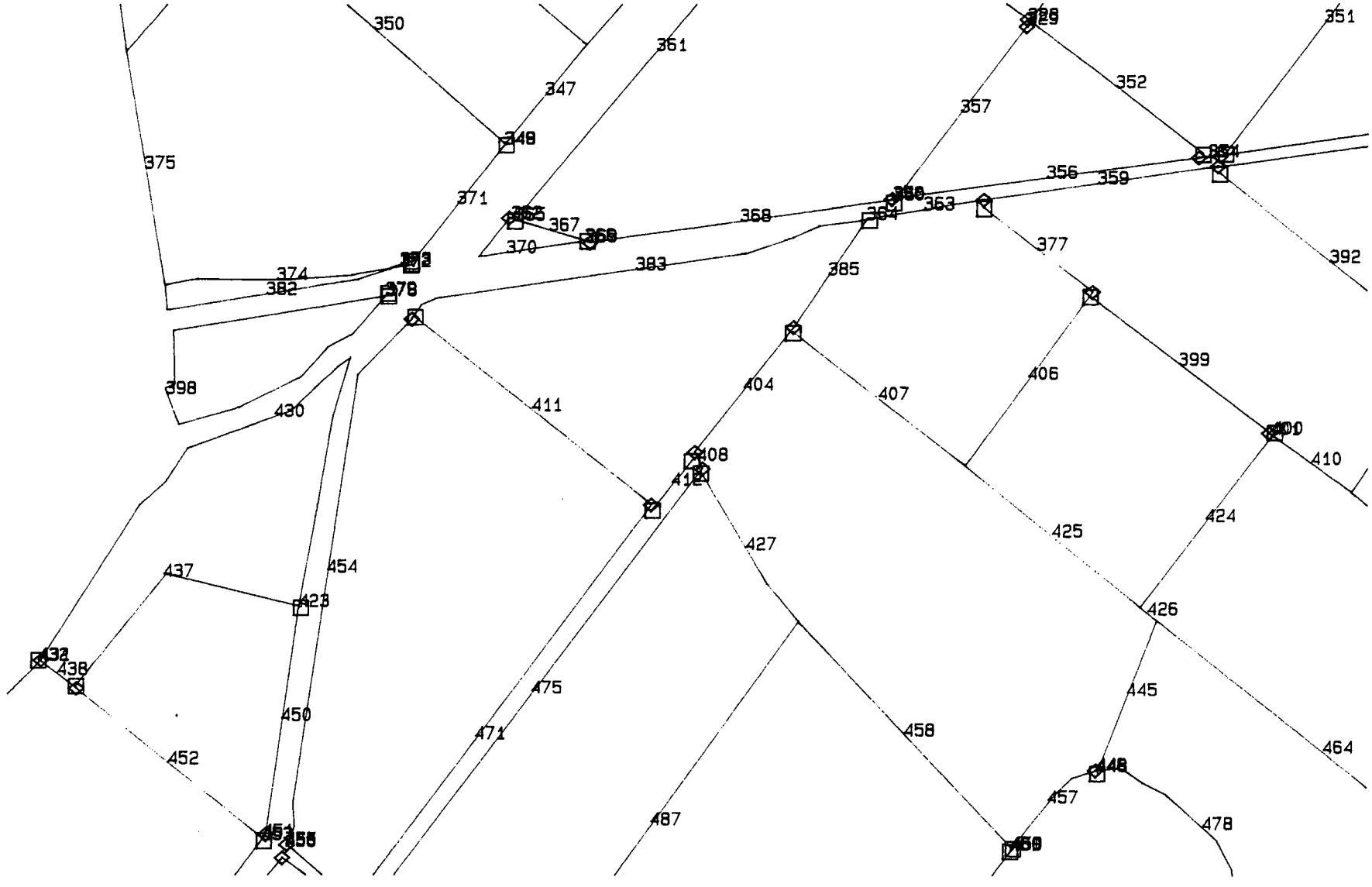
0	-.40	
0	-.25	
0	-.10	
564	.05	**********
27	.20	**
6	.35	*
5	.50	
4	.65	
6	.80	*
1	.95	
0	1.10	
0	1.25	
0	1.40	
0	1.55	
2	1.70	
0	1.85	
0	2.00	
1	2.15	
0	2.30	
0	2.45	
0	2.60	

T009C2W1

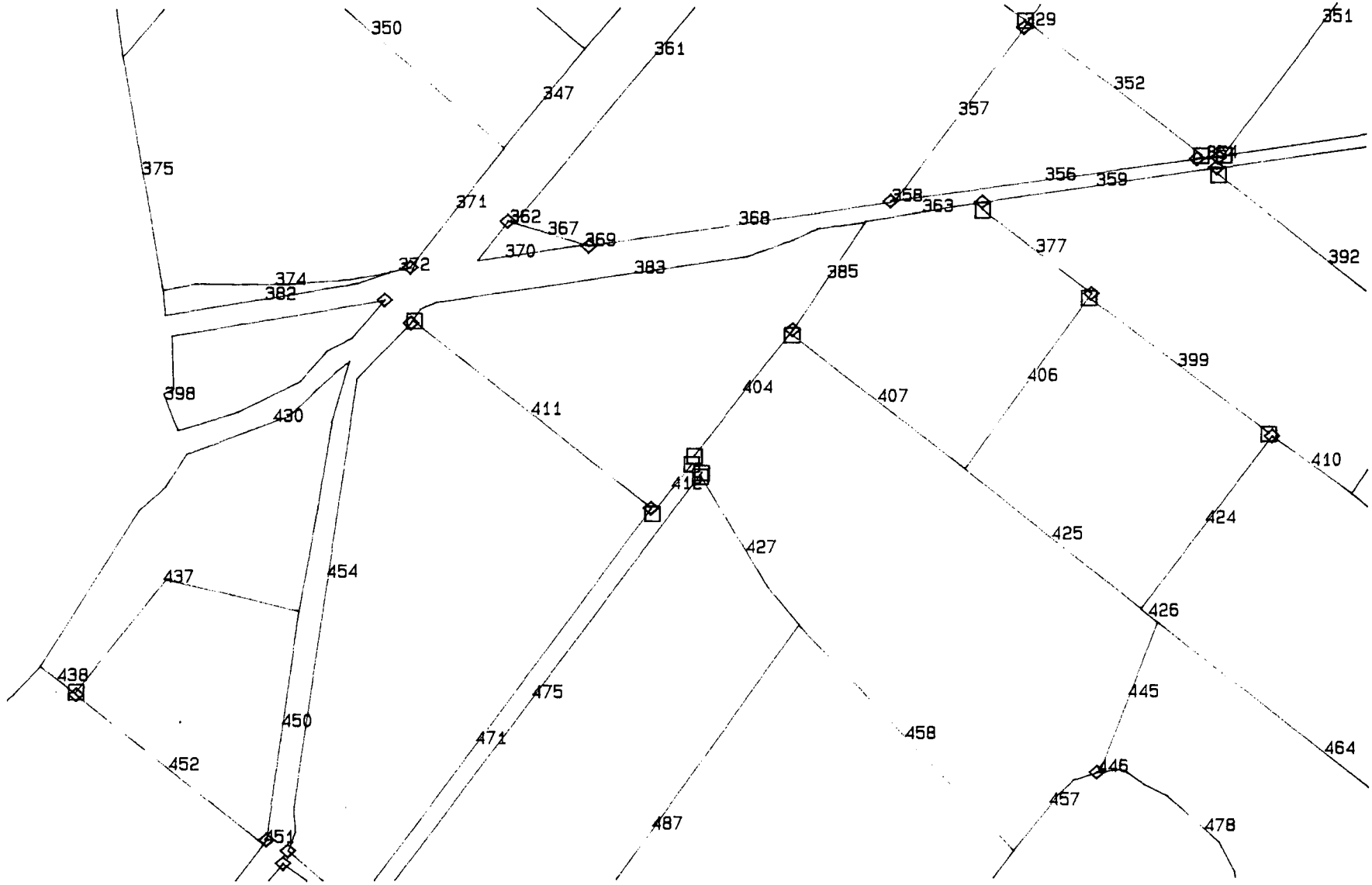
I....+....I....+....I....+....I....+....I....+....I....+....I
0 120 240 360 480 600
HISTOGRAM FREQUENCY



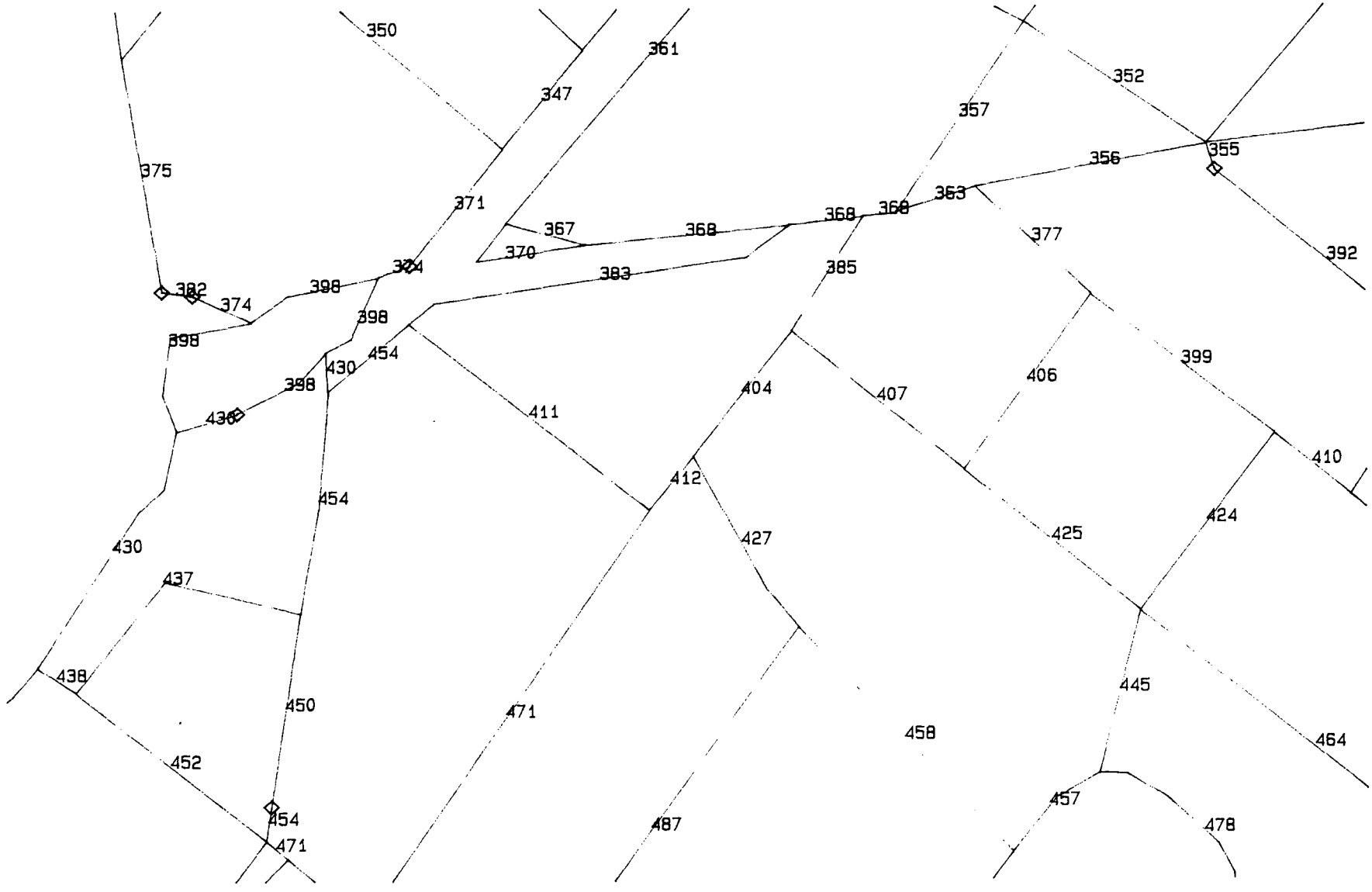
5.83 Subset of the original coverage showing line numbers and node errors



5.84 Subset of T009C03 coverage showing line numbers and node errors

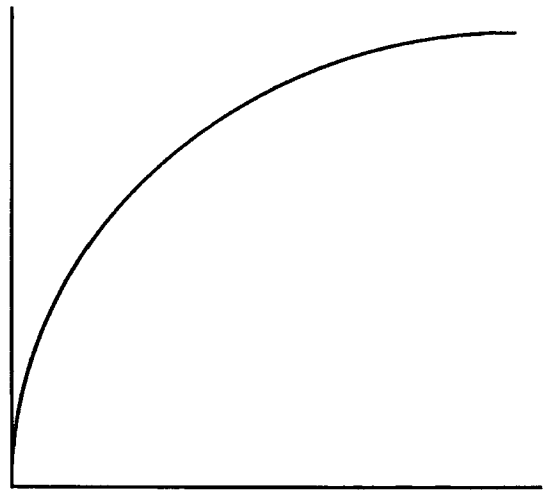


5.85 Subset of T009C2 coverage showing line numbers and node errors

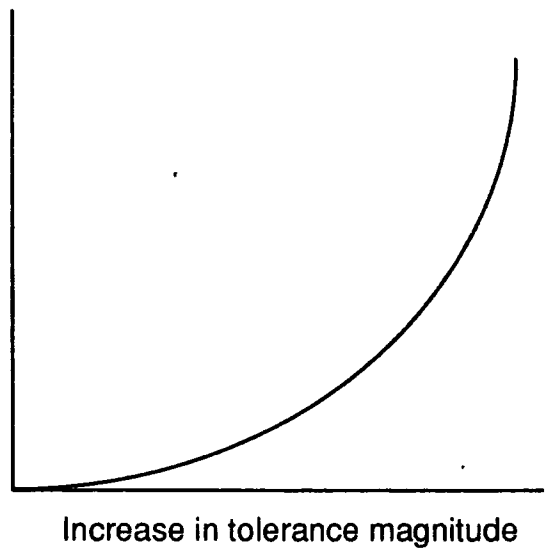


5.86 Subset of T009C2W1 coverage showing line numbers and node errors

5.87 Perspectives on geometric change as tolerance values are increased



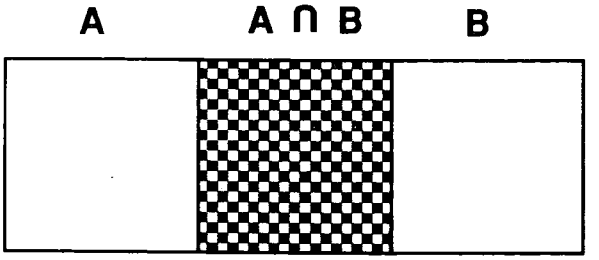
Optimisitc curve for geometric change following tolerancing



Pessimisitc curve for geometric change following tolerancing

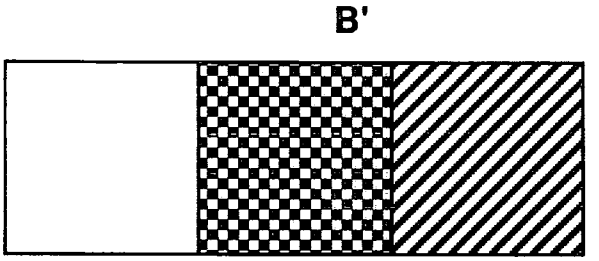
Figures for Chapter 6

6.1 The nature of set theory adoption in ARC/INFO



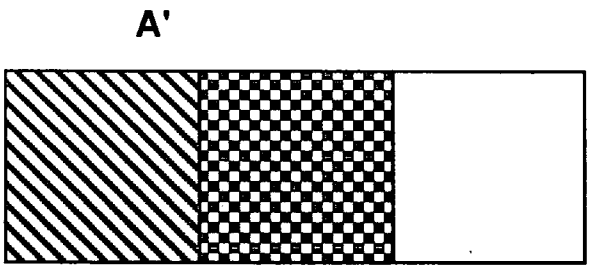
The intersection of A and B

Acceptable



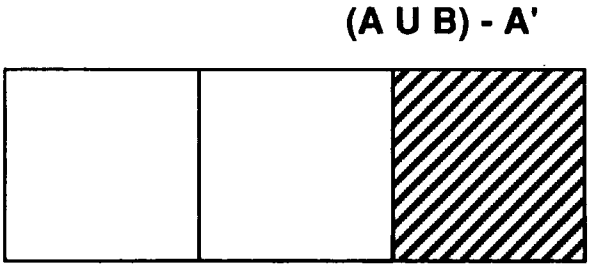
The identity of B

Acceptable



The identity of A

Acceptable

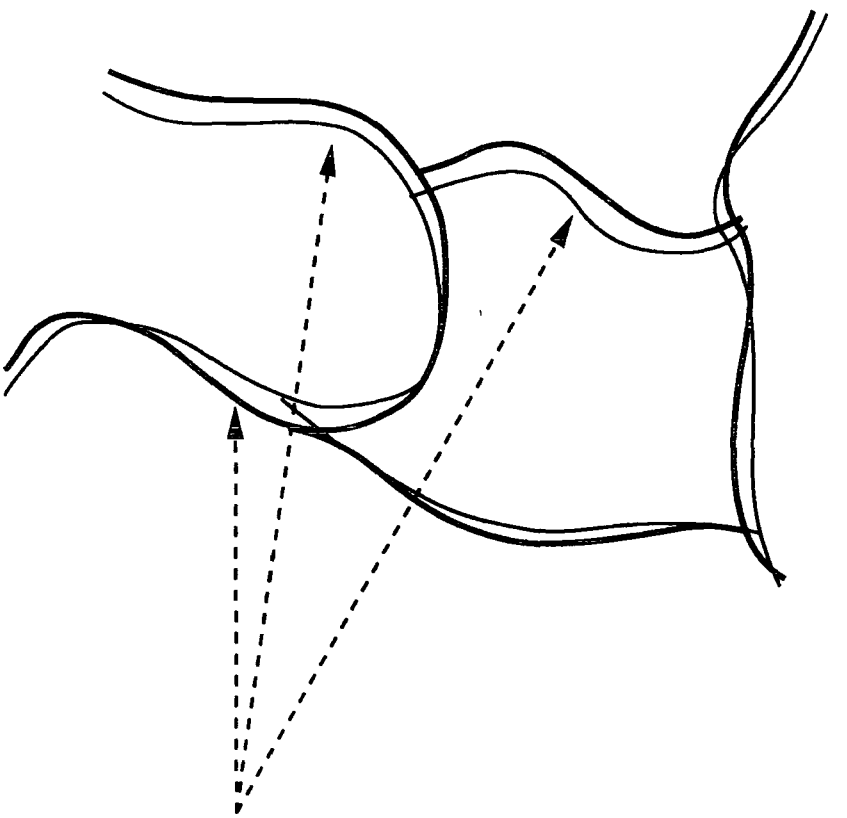


The union of A and B minus the identity of A

Unacceptable

Thus in terms of its geometric overlay options, the system favours accumulation of information rather than selective removal

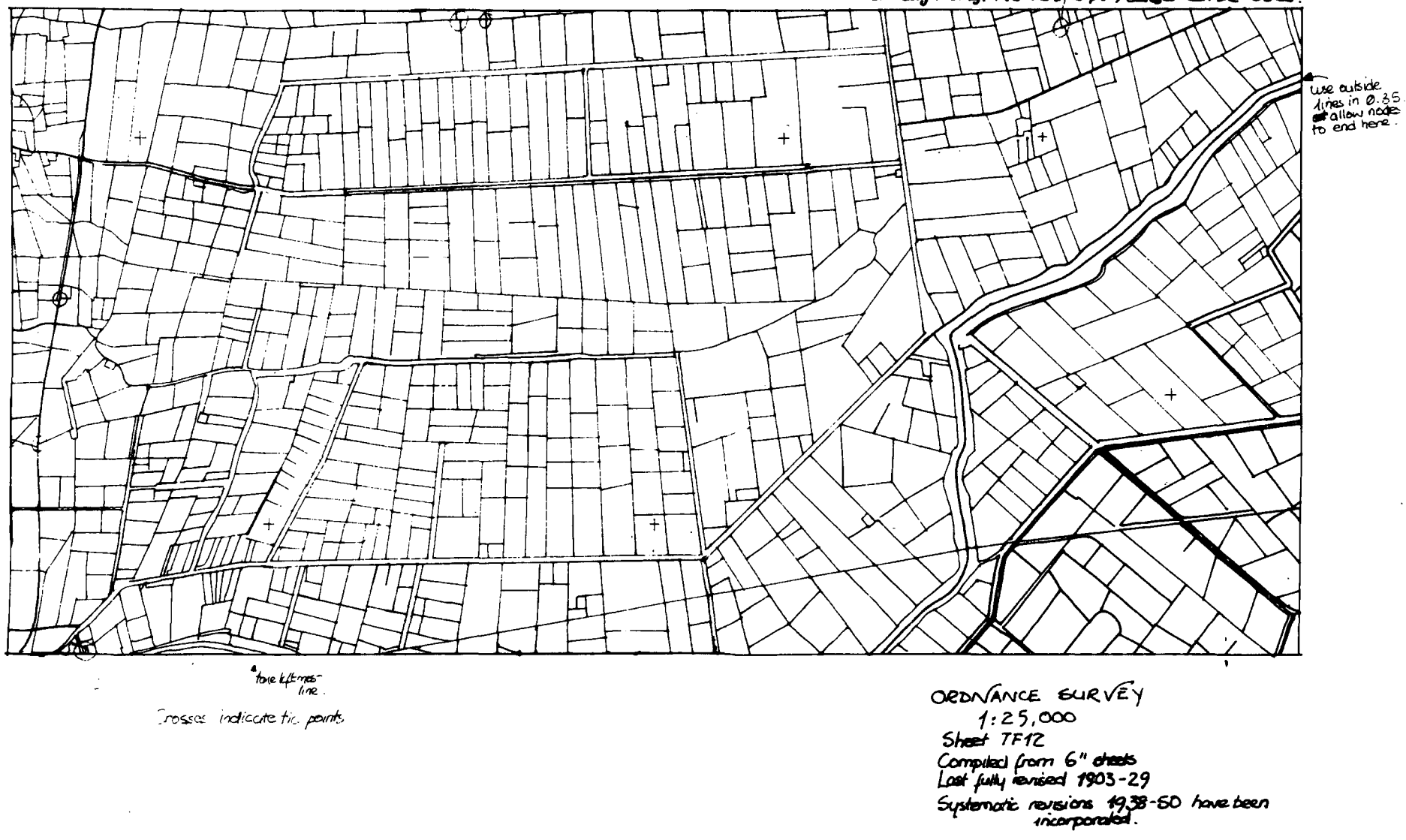
6.2 Imperfect linear matching in overlay due to the variability of source data



Zones of mis-match

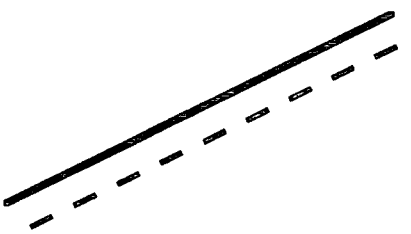
Map as drawn from souce A

Map as drawn from source B

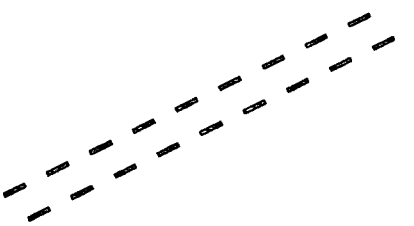


6.3 A tracing of the 1:25,000 Ordnance Survey map used as input showing feature generalisation

6.4 The transformation of dual lines into single entities

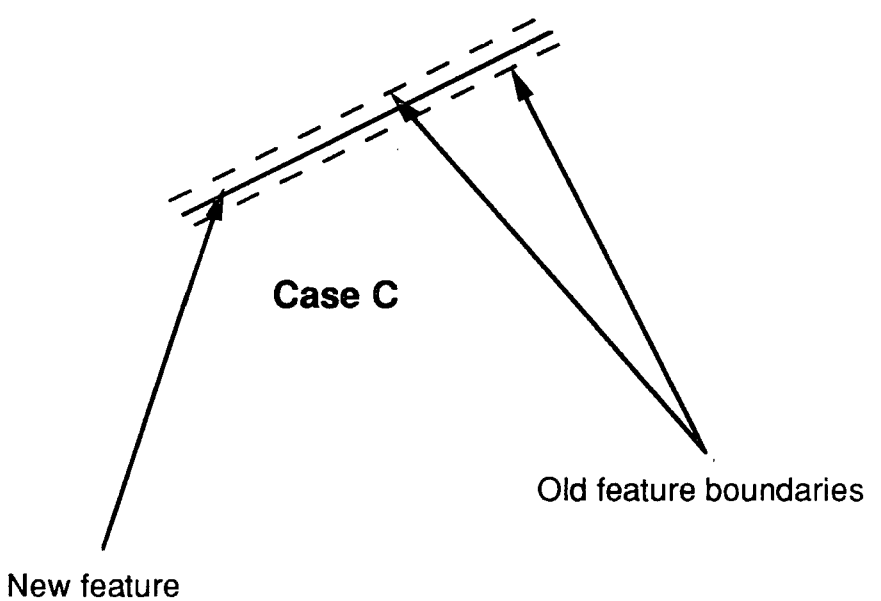


Case A

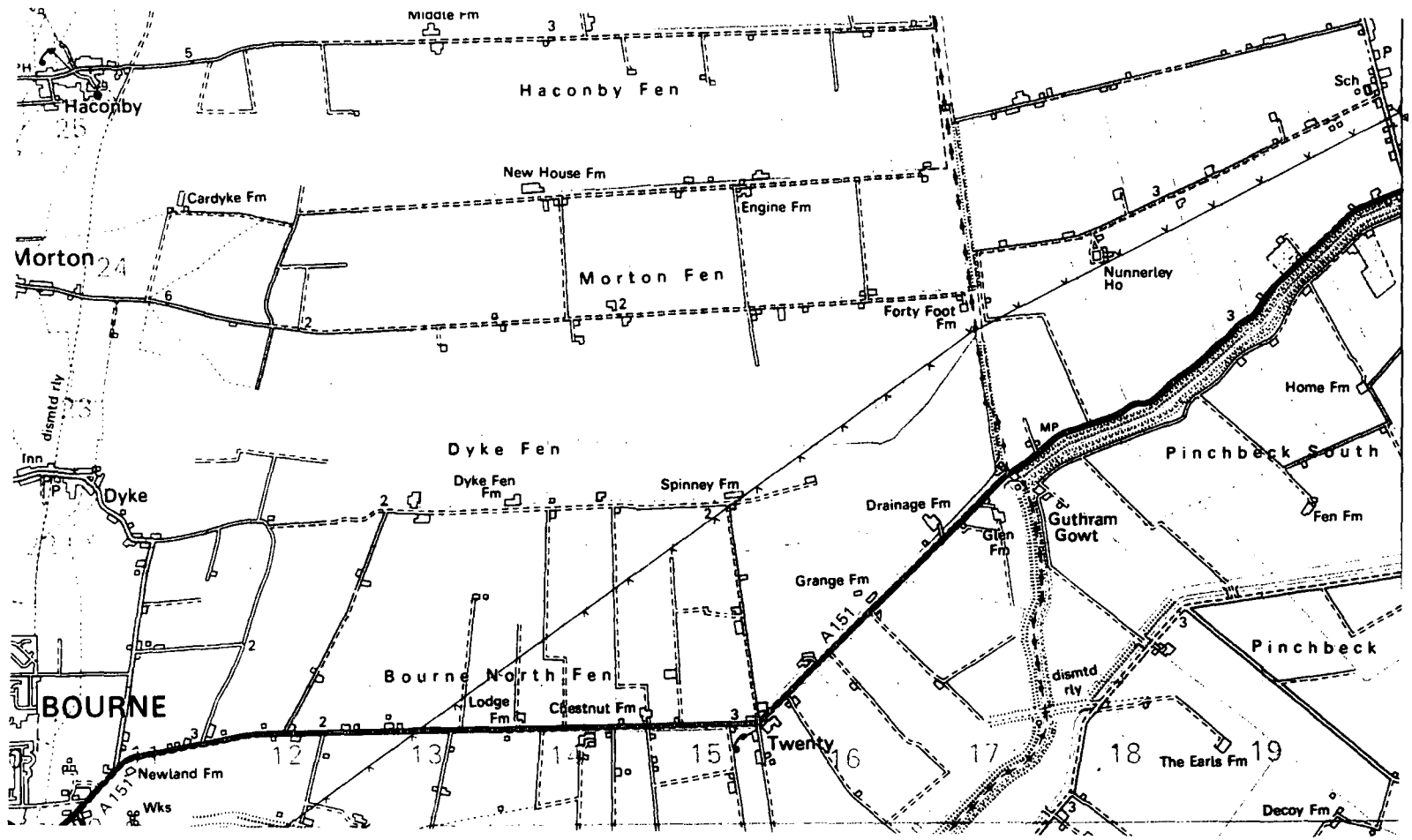


Case B

The depiction of linear features on Ordnance Survey maps using dual lines (such as in case A or B) was translated into a single fine line (case C) when the data was transcribed from the original map sheets

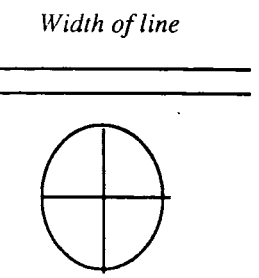


6.5 A section of the 1:50,000 Ordnance Survey map used as the basis for coverage input



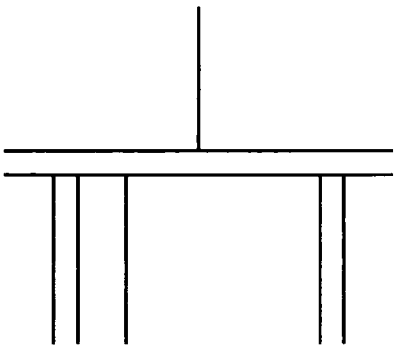
Copyright Ordnance Survey

.6. Problems of digitising lines represented as areas in source locument.



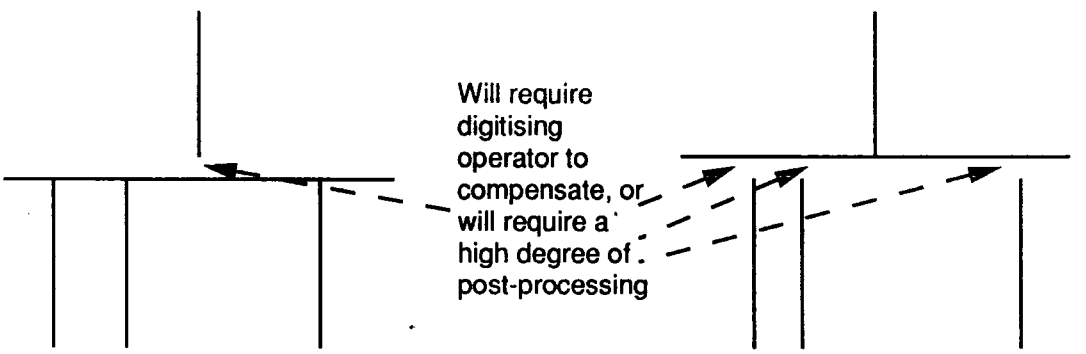
Crosshair

If the cross hair is actually smaller than the width of the line being dealt with then the user will have to decide whic of the line boundaries are traced.



Original coverage consisting of thickly drawn lines

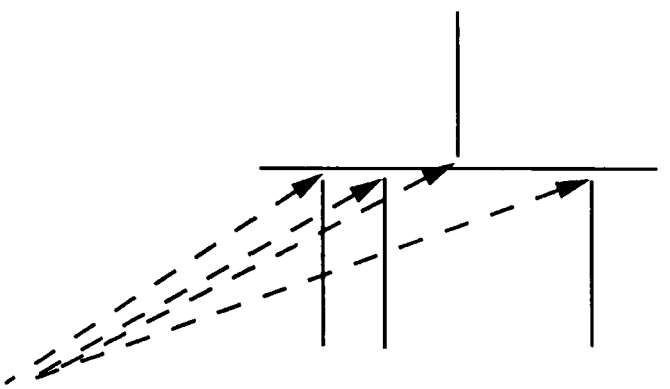
Following either the top or the bottom part of the line and/or the left or the right part of the line would produce results such as the following two cases.



Digitise the bottom line and the left boundaries

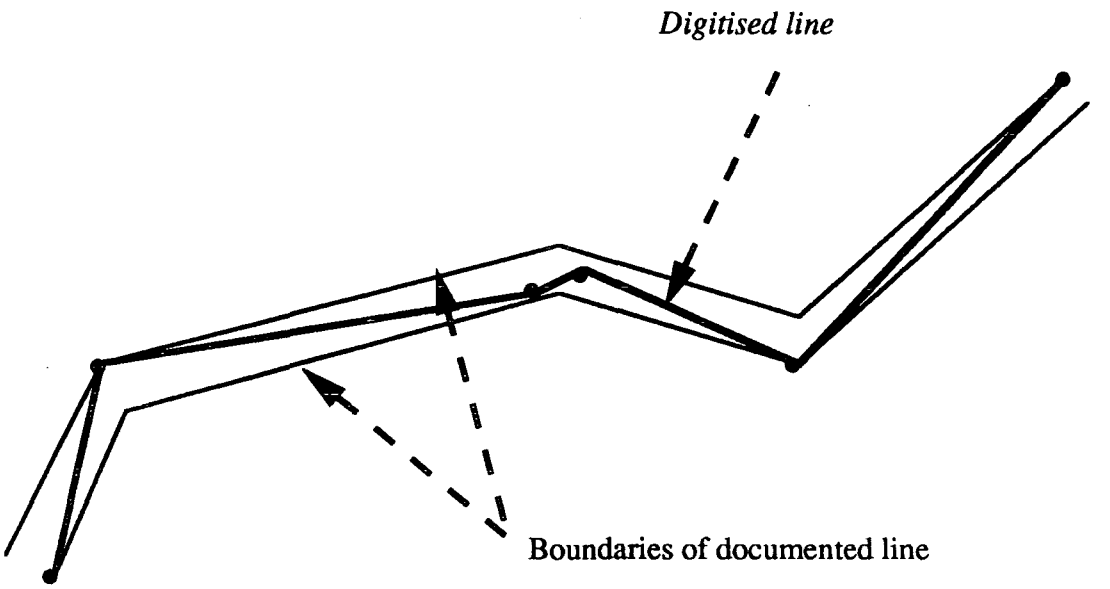
Digitise the top line and the right boundaries

Alternatively, the digitising operator may have to approximate a centreline which could result in the following scenario;



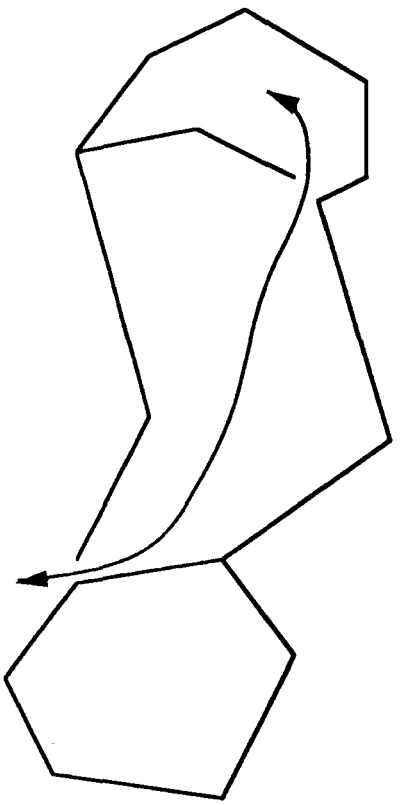
Will require compensation by digitisng operator 'on the fly' or during post-processing

6.7. Problems of the 'bouncing' line during digitising.



The thin line of the digitising cross-hair is finer than the line that is being digitised and as a consequence, the digitised line 'bounces' off the walls of the documented line

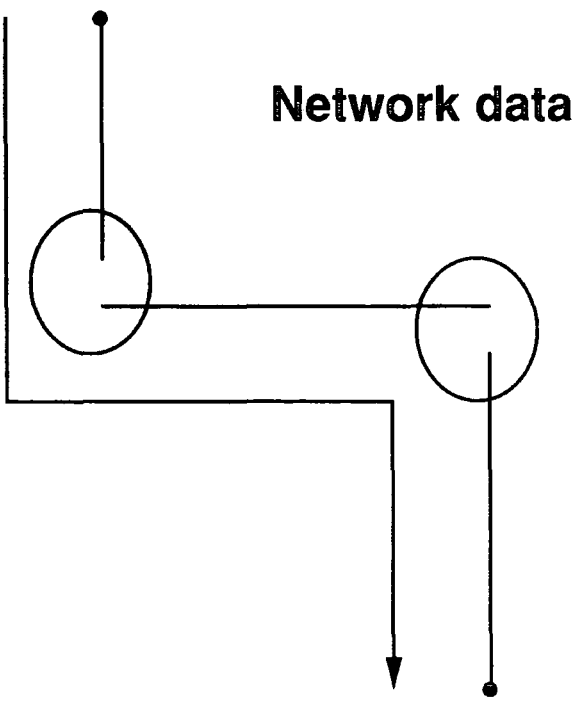
6.8. Problems of topology as a consequence of digitising errors



Polygon data

Areas 'flow ' into each other and the software sees the entire extent as one area, and does not allow topology to be built

Start point

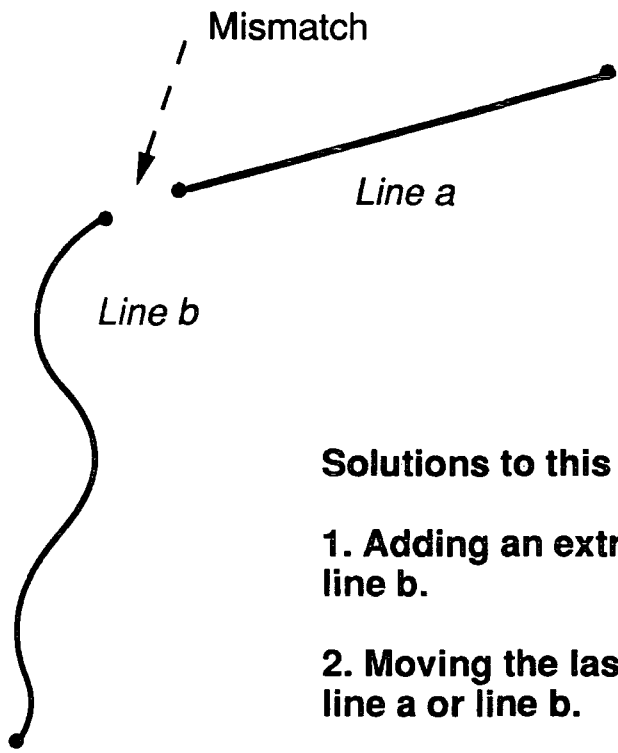


Network data

Gaps do not allow a flow to be established between the start and endpoints, and therefore no analysis can take place.

End point

6.9 Options for manual editing of geometric discrepancies



Solutions to this mismatch include;

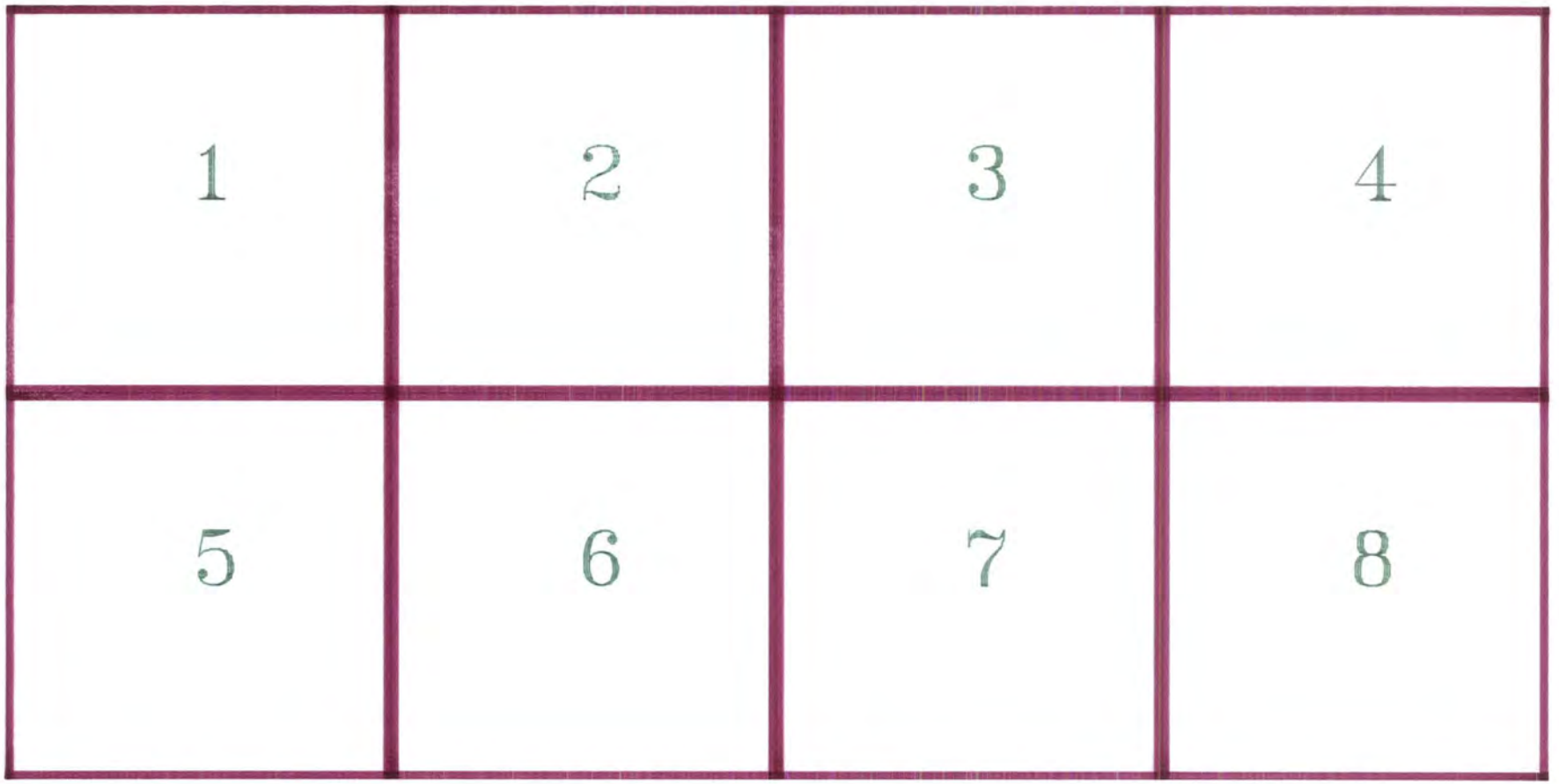
1. Adding an extra vertex to either line a or line b.

2. Moving the last vertex attached to either line a or line b.

3. Deleting the current version and redrawing it.

4. Joining the two arcs by adding an extra arc.

6.10. Diagram showing the division of the full map into eight sub-sections



6.11

Map of the Fenlands digitised from the 1:10560 scale map

510000 325000

520000 325000

510000 320000

520000 320000

6.12

Map of the Fenlands digitised from the 1:10560 scale map

510000 325000

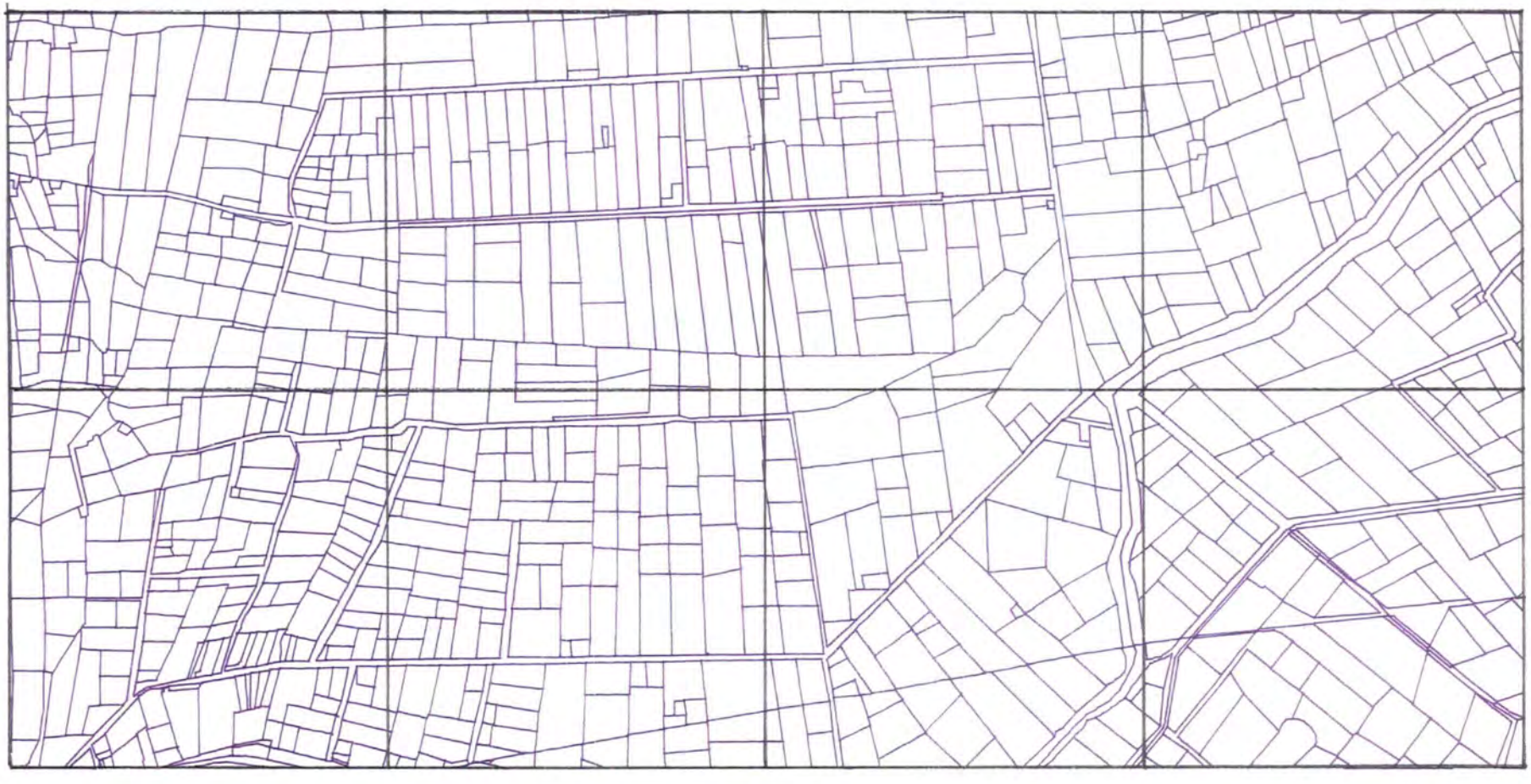
520000 325000

510000 320000

520000 320000

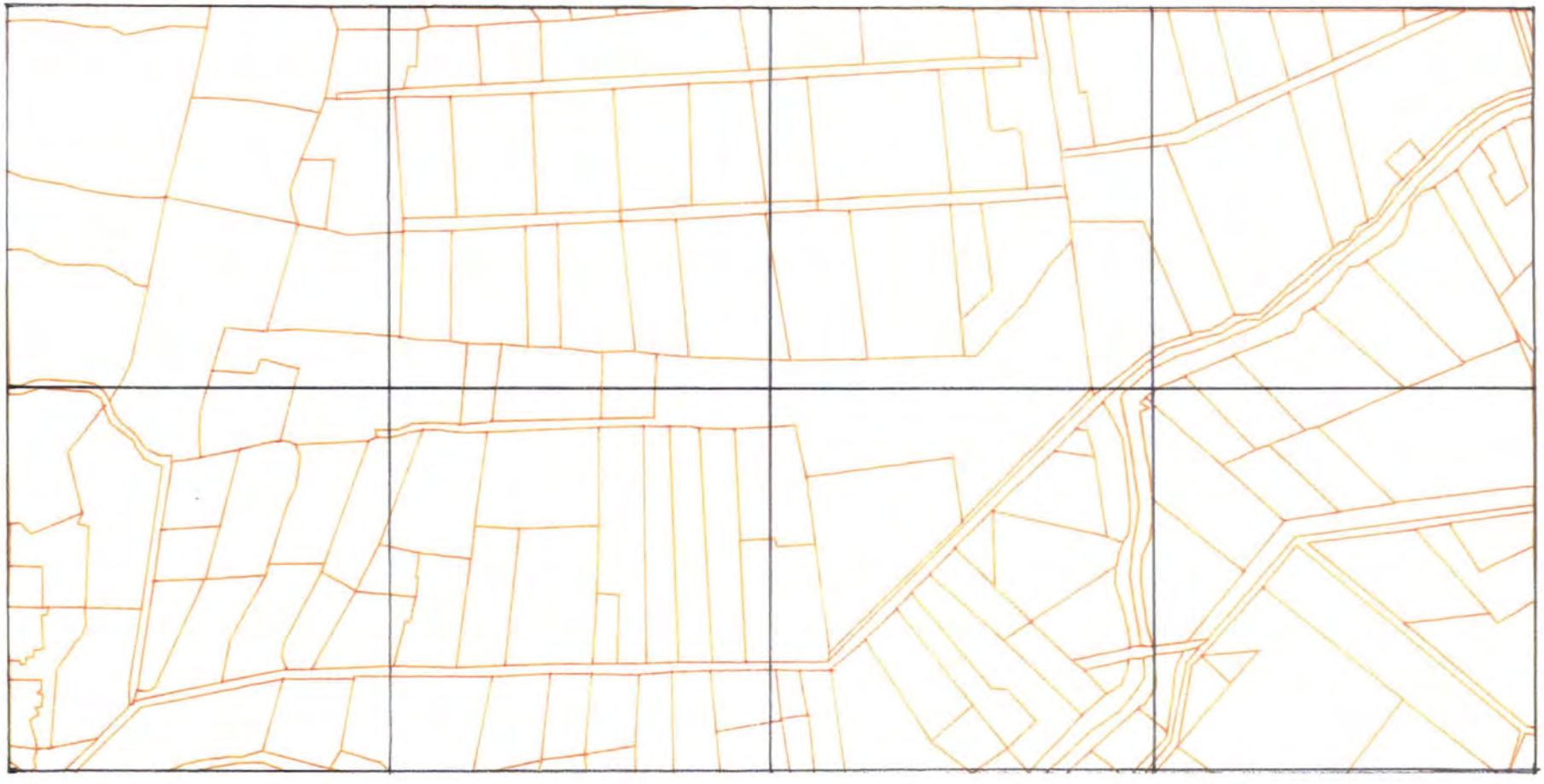
6.13 Map of the Fenlands digitised from the 1:25000 scale map

510000 325000 520000 325000



510000 320000 520000 320000

6.14 Map of the Fenlands digitised from the 1:50000 scale map

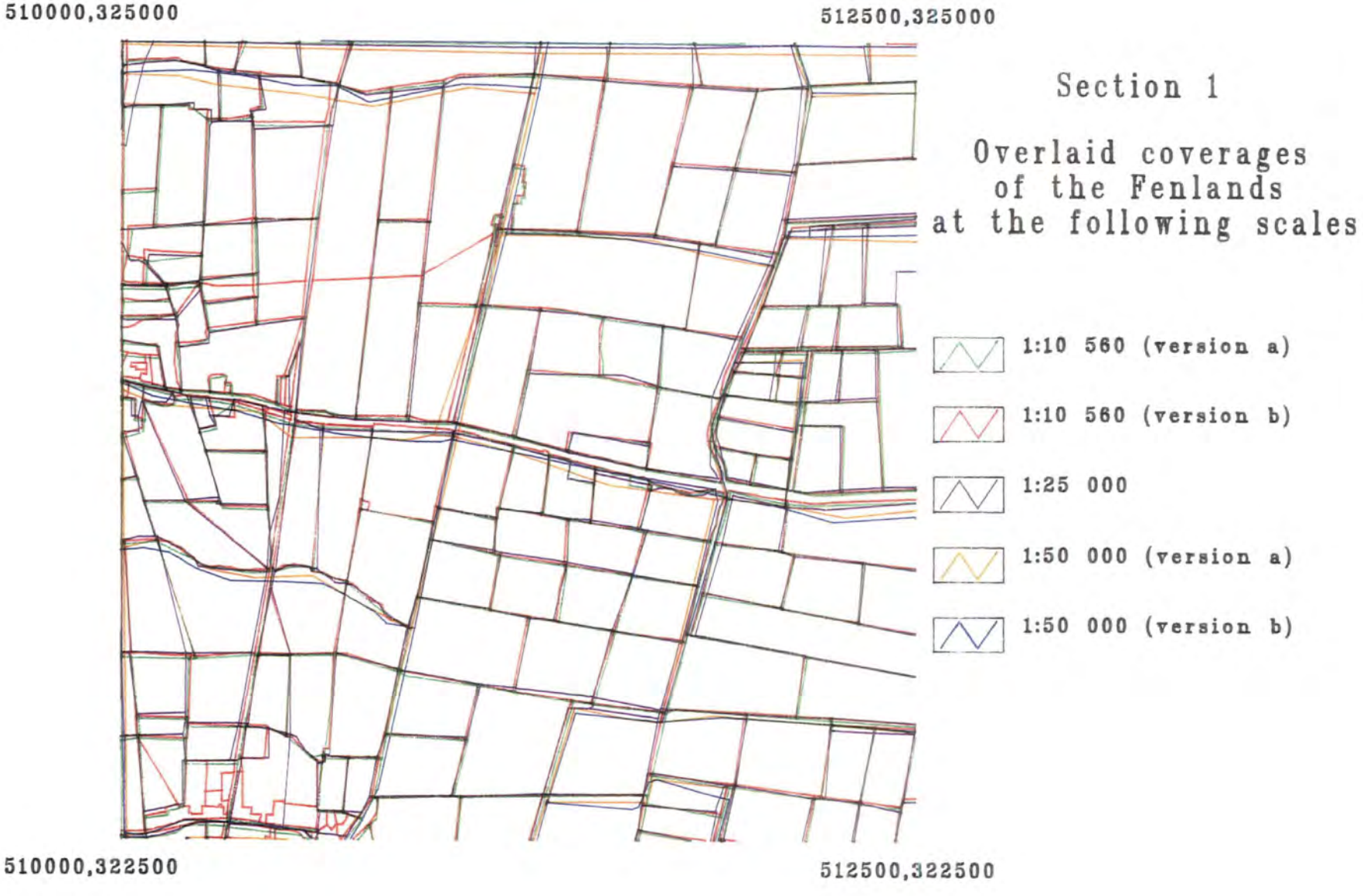


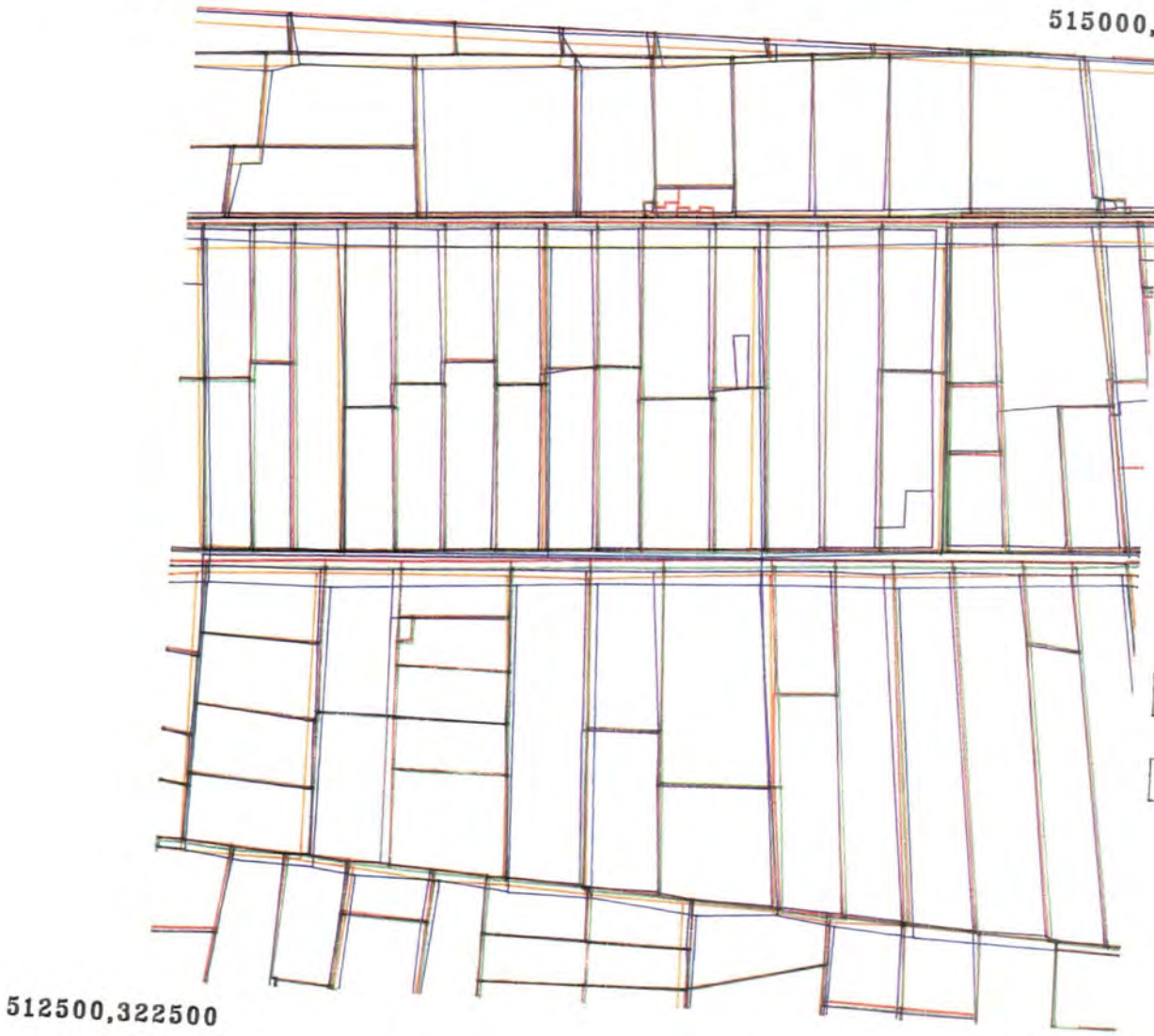
6.15

Map of the Fenlands digitised from the 1:50000 scale map

6.16 Overlay of the two 1:50,000 scale coverages

6.17


510000,325000
512500,325000
Section 1
Overlaid coverages
of the Fenlands
at the following scales
1:10 560 (version a)
1:10 560 (version b)
1:25 000
1:50 000 (version a)
1:50 000 (version b)
510000,322500
512500,322500



6.18
Section 2

Overlaid coverages
of the Fenlands
at the following scales

1:10 560 (version a)

1:10 560 (version b)

1:25 000

1:50 000 (version a)

1:50 000 (version b)

515000,325000

517500,325000

515000,322500

517500,322500

6.19
Section 3

Overlaid coverages
of the Fenlands
at the following scales

1:10 560 (version a)

1:10 560 (version b)

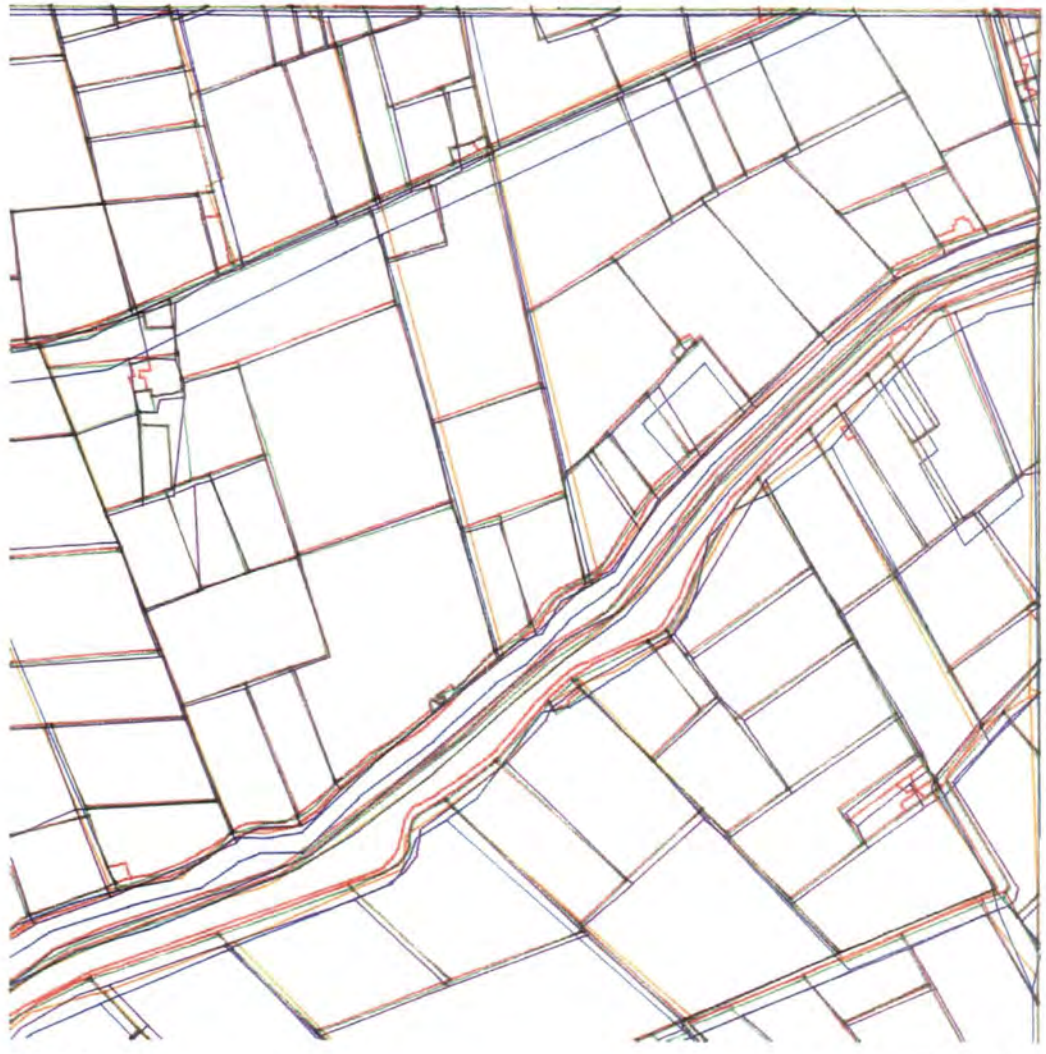
1:25 000

1:50 000 (version a)

1:50 000 (version b)

517500,325000

520000,325000



517500,322500

520000,322500

6.20
Section 4

Overlaid coverages
of the Fenlands
at the following scales

- 1:10 560 (version a)
- 1:10 560 (version b)
- 1:25 000
- 1:50 000 (version a)
- 1:50 000 (version b)

510000,322500

512500,322500

510000,320000

512500,320000

6.21

Section 5

Overlaid coverages of the Fenlands at the following scales

1:10 560 (version a)

1:10 560 (version b)

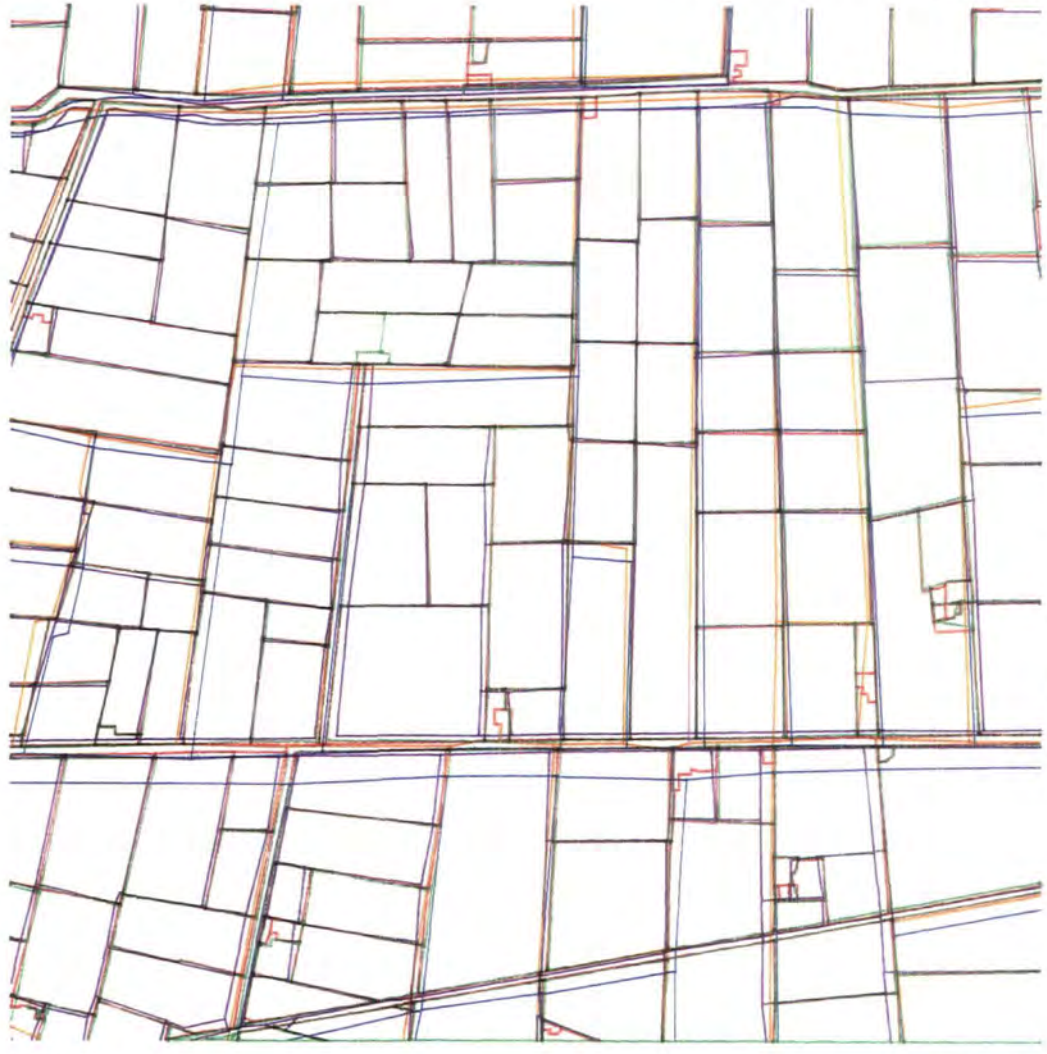
1:25 000

1:50 000 (version a)

1:50 000 (version b)

512500,322500

515000,322500



6.22
Section 6

Overlaid coverages
of the Fenlands
at the following scales

1:10 560 (version a)

1:10 560 (version b)

1:25 000

1:50 000 (version a)

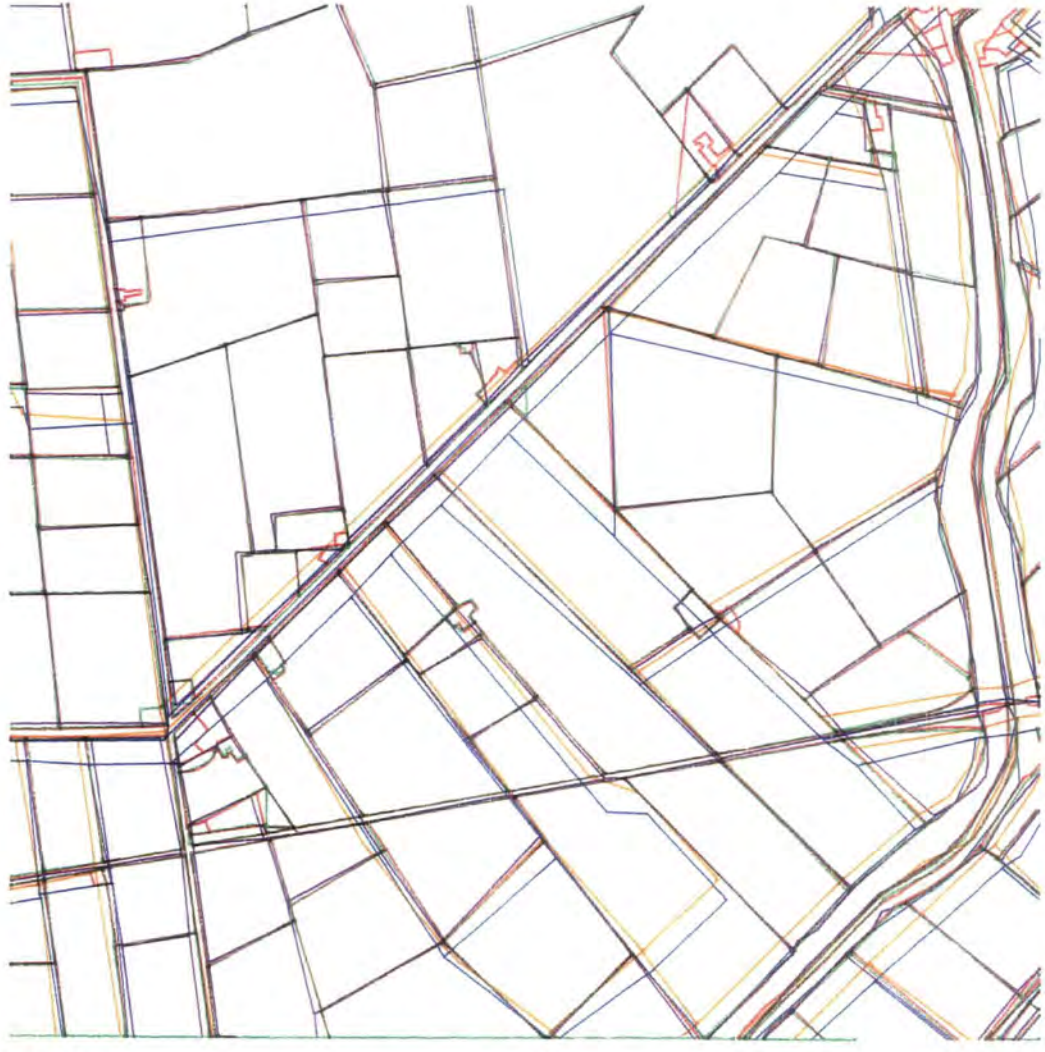
1:50 000 (version b)

512500,320000

515000,320000

515000,322500

517500,322500



515000,320000

517500,320000

6.23

Section 7

Overlaid coverages of the Fenlands at the following scales

1:10 560 (version a)

1:10 560 (version b)

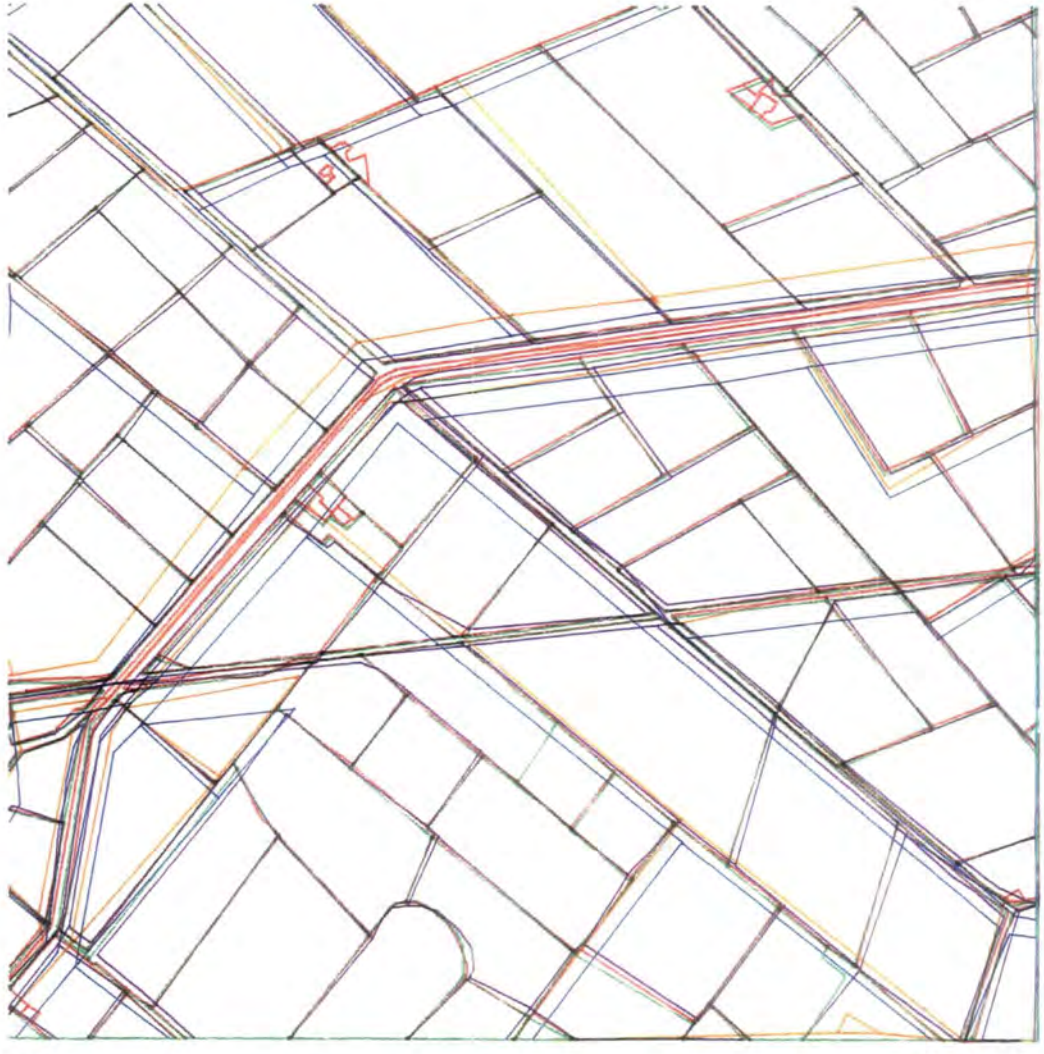
1:25 000

1:50 000 (version a)

1:50 000 (version b)

517500,322500

520000,322500



6.24
Section 8

Overlaid coverages
of the Fenlands
at the following scales

1: 10 560 (version a)

1: 10 560 (version b)

1: 25 000

1: 50 000 (version a)

1: 50 000 (version b)

517500,320000

520000,320000

6.25 A frequency histogram of polygon size for the 1:50,000b coverage

```
   Count    Midpoint     One symbol equals approximately  1.20 occurrences

      55       50162  ***********************************************
      40      150476  **********************************
      30      250790  ****************************
      24      351104  ********************
      14      451418  ************
      13      551732  ***********
       5      652046  ****
       1      752360  *
       2      852674  **
       2      952988  **
       1     1053302  *
       1     1153616  *
       1     1253930  *
       0     1354244
       0     1454558
       0     1554872
       0     1655186
       0     1755500
       0     1855814
       0     1956128
       1     2056442  *
                      I....+....I....+....I....+....I....+....I....+....I....+....I
                      0         12         24         36         48         60
                                Histogram frequency
```

Mean	262775.459	Std err	19494.874	Median	196712.257
Mode	6.731	Std dev	268718.289	Variance	7.221E+10
Kurtosis	12.238	S E Kurt	.351	Skewness	2.615
S E Skew	.176	Range	2106589.77	Minimum	6.731
Maximum	2106596.50	Sum	49927337.2		

Valid cases	190	Missing cases	1

6.26 A frequency histogram of polygon size for the overlay of the two 1:50,000 coverages

```
  COUNT    MIDPOINT     ONE SYMBOL EQUALS APPROXIMATELY 16.00 OCCURRENCES

    753       47826   **************************************************
     44      143494   ***
     25      239160   **
     24      334826   **
     10      430492   *
     13      526158   *
      3      621824
      2      717490
      2      813156
      2      908822
      1     1004488
      1     1100154
      1     1195820
      0     1291486
      0     1387152
      0     1482818
      0     1578484
      0     1674150
      0     1769816
      0     1865482
      1     1961148

                    I....+....I....+....I....+....I....+....I....+....I....+....I
                    0        160        320        480        640        800
                              HISTOGRAM FREQUENCY

MEAN          56816.829      STD ERR     5123.505      MEDIAN      5938.540
MODE              .850       STD DEV   152160.323      VARIANCE  2.315E+10
KURTOSIS        42.487       S E KURT       .164      SKEWNESS       5.280
S E SKEW          .082       RANGE     2008973.90      MINIMUM         .850
MAXIMUM    2008974.75        SUM       50112443.1
```

6.27 A frequency distribution of polygon size in the 1:10,560b coverage

```
  Count    Midpoint     One symbol equals approximately 16.00 occurrences
    713       24038   ********************************************
    293       72115   ******************
     70      120192   ****
     25      168269   **
      9      216346   *
      5      264423
      2      312500
      0      360577
      1      408654
      0      456731
      0      504808
      0      552885
      0      600962
      0      649039
      0      697116
      0      745193
      0      793270
      0      841347
      0      889424
      0      937501
      1      985578

                    I....+....I....+....I....+....I....+....I....+....I....+....I
                    0        160        320        480        640        800
                              Histogram frequency
```

Mean	44709.736	Std err	1620.864	Median	34985.809
Mode	6.352	Std dev	54220.264	Variance	2939836988
Kurtosis	93.077	S E Kurt	.146	Skewness	6.369
S E Skew	.073	Range	1009604.27	Minimum	6.352
Maximum	1009610.63	Sum	50030194.4		

Valid cases	1119	Missing cases	1

6.28

A frequency distribution of polygon size in the overalid coverage consisting of two versions of the 1:10560 Ordnance Survey maps

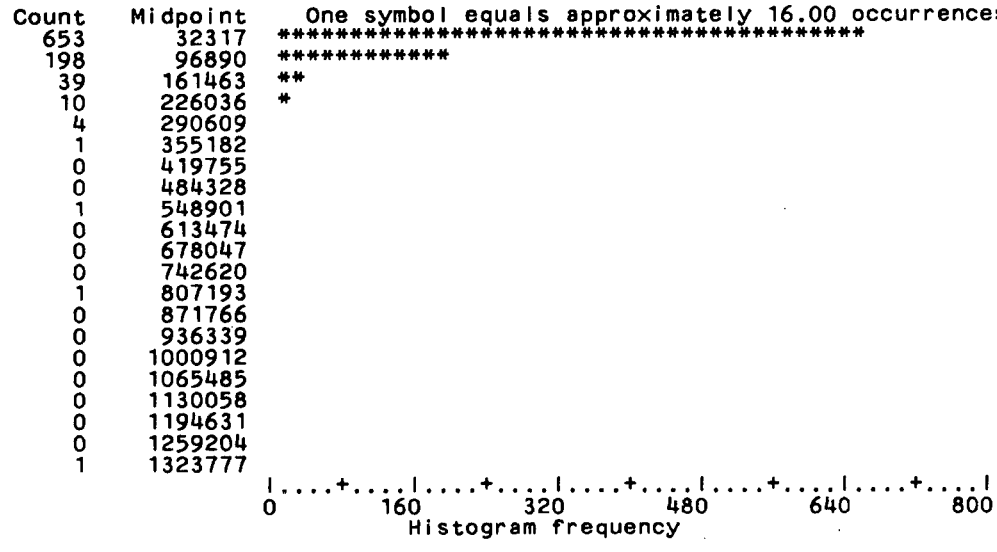
COUNT	MIDPOINT	ONE SYMBOL EQUALS APPROX. 160.00 OCCURRENCES
4327	15009	**********
319	45027	**
158	75045	*
53	105063	
29	135081	
17	165099	
9	195117	
2	225135	
4	255153	
4	285171	
0	315189	
0	345207	
0	375225	
0	405243	
0	435261	
0	465279	
0	495297	
0	525315	
0	555333	
0	585351	
1	615369	

I....+....I....+....I....+....I....+....I....+....I....+....I
0 1600 3200 4800 6400 8000
HISTOGRAM FREQUENCY

MEAN	10165.601	STD ERR	406.949	MEDIAN	515.560
MODE	.671	STD DEV	28553.223	VARIANCE	815286551
KURTOSIS	65.454	S E KURT	.070	SKEWNESS	5.819
S E SKEW	.035	RANGE	630377.644	MINIMUM	.231
MAXIMUM	630377.875	SUM	50045253.4		

6.29

A FREQUENCY HISTOGRAM SHOWING THE DISTRIBUTION OF POLYGON SIZE FOR THE 1:25,000 COVERAGE.

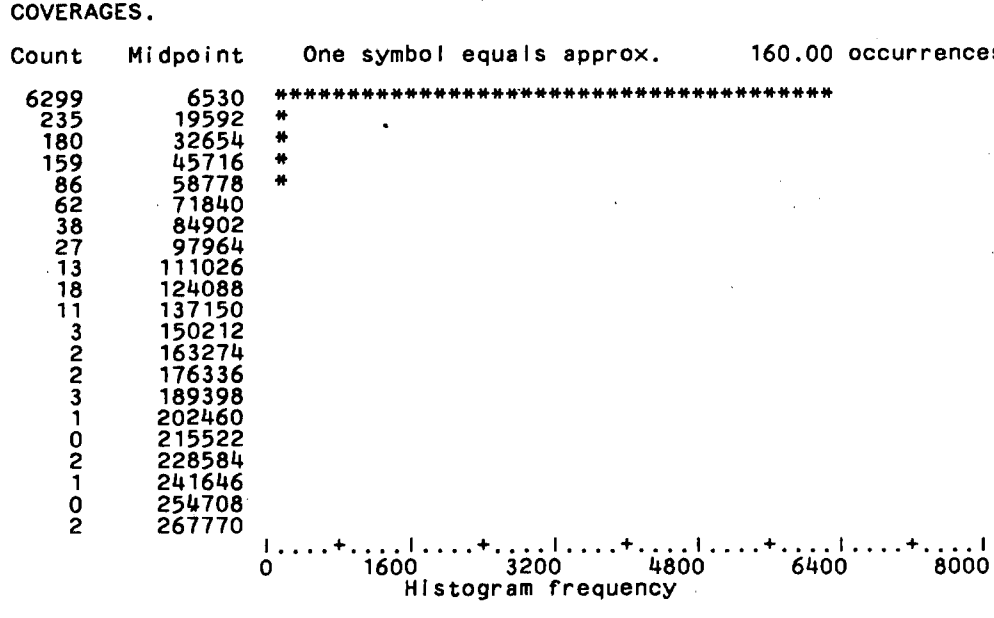


Mean	55019.966	Std err	2277.307	Median	42415.600
Mode	32.909	Std dev	68622.183	Variance	4709004062
Kurtosis	161.044	S E Kurt	.162	Skewness	9.911
S E Skew	.081	Range	1356027.34	Minimum	32.909
Maximum	1356060.25	Sum	49958129.3		

Valid cases 908 Missing cases 1

6.30

A FREQUENCY HISTOGRAM SHOWING THE DISTRIBUTION OF POLYGON SIZE FOLLOWING AN OVERLAY OF THE 1:50,000B, 1:25,000 AND 1:10,560 COVERAGES.



Mean	7023.627	Std err	230.732	Median	713.156
Mode	1.300	Std dev	19501.955	Variance	380326231
Kurtosis	35.257	S E Kurt	.058	Skewness	5.035
S E Skew	.029	Range	274299.396	Minimum	.260
Maximum	274299.656	Sum	50176792.4		
VALID CASES	7144	MISSING CASES	0		

6.3 f A frequency histogram of polygon size following an overlay of the 1:10,560b coverage and the 1:25,000 coverage

Count	Midpoint	One symbol equals approx. 160.00 occurrences
4134	12677	**********
306	38043	**
176	63409	*
82	88775	*
34	114141	
18	139507	
10	164873	
7	190239	
4	215605	
2	240971	
2	266337	
1	291703	
0	317069	
0	342435	
0	367801	
0	393167	
0	418533	
0	443899	
0	469265	
0	494631	
1	519997	

```
I....+....I....+....I....+....I....+....I....+....I....+....I
0       1600       3200       4800       6400       8000
          Histogram frequency
```

Mean	10488.881	Std err	385.894	Median	984.460
Mode	1.300	Std dev	26671.355	Variance	711361198
Kurtosis	50.342	S E Kurt	.071	Skewness	5.240
S E Skew	.035	Range	532673.053	Minimum	.260
Maximum	532673.313	Sum	50105386.7		

6.32 A frequency histogram of the two overlays involving the 1:50,000b coverage and coverages of other scales

A FREQUENCY DISTRIBUTION OF POLYGON SIZE, FOLLOWING AN OVERLAY OF THE 1:10,560B COVERAGE AND THE 1:50,000B COVERAGE

Count	Midpoint	One symbol equals approximately 40.00 occurrences
1665	6750	**********
252	20255	******
176	33760	****
144	47265	****
103	60770	***
68	74275	**
37	87780	*
32	101285	*
10	114790	
20	128295	*
13	141800	
3	155305	
2	168810	
2	182315	
3	195820	
2	209325	
0	222830	
2	236335	
2	249840	
1	263345	
2	276850	

```
I....+....I....+....I....+....I....+....I....+....I....+....I
0         400         800        1200        1600        2000
          Histogram frequency
```

Mean	19740.935	Std err	630.788	Median	5490.389
Mode	1.038	Std dev	31784.437	Variance	1010250451
Kurtosis	12.556	S E Kurt	.097	Skewness	2.967
S E Skew	.049	Range	283598.400	Minimum	1.038
Maximum	283599.438	Sum	50122233.1		

Valid cases	2539	Missing cases	1

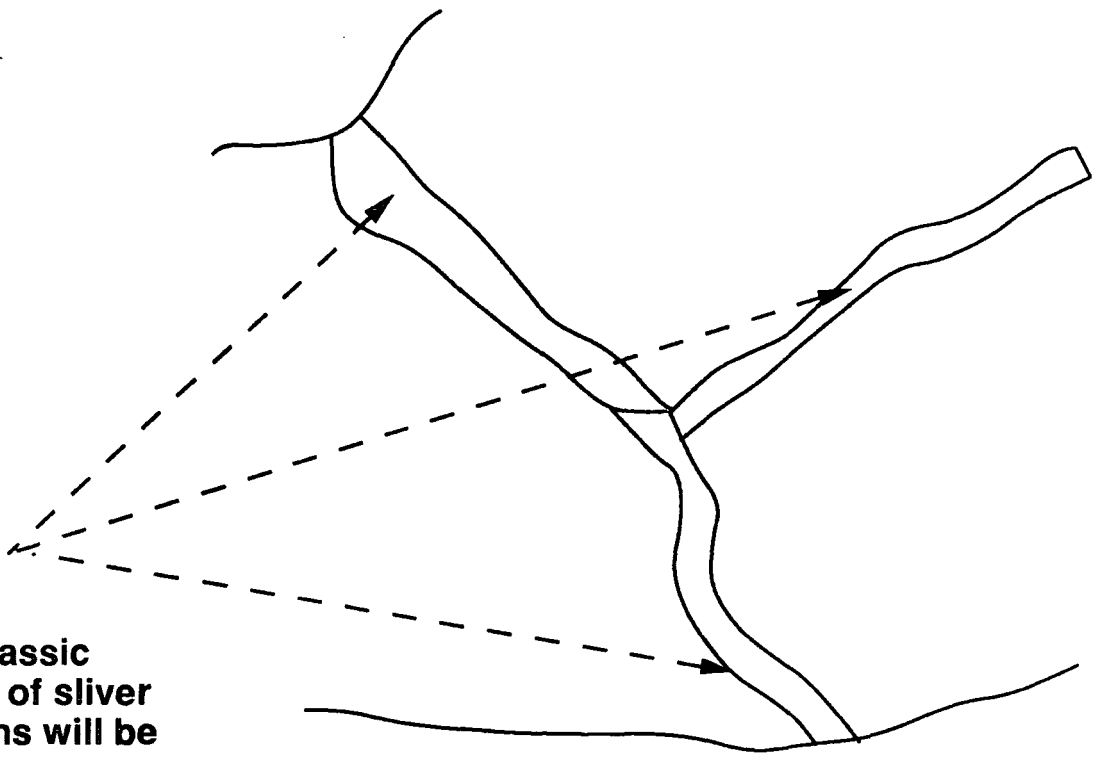
A frequency histogram of polygon size following an overlay of the 1:25,000 coverage and the 1:50,00b coverage

COUNT	MIDPOINT	ONE SYMBOL EQUALS APPROXIMATELY 30.00 OCCURRENCES
1341	9141	**********
232	27425	********
136	45709	********
108	63993	****
67	82277	**
41	100561	*
29	118845	*
16	137129	*
4	155413	
2	173697	
5	191981	
4	210265	
2	228549	
1	246833	
1	265117	
1	283401	
0	301685	
0	319969	
0	338253	
0	356537	
1	374821	

```
I....+....I....+....I....+....I....+....I....+....I....+....I
0         300         600         900        1200        1500
          HISTOGRAM FREQUENCY
```

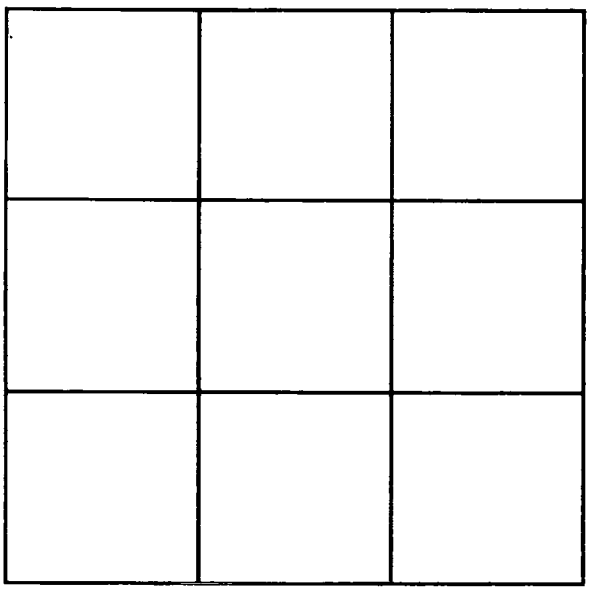
MEAN	23721.759	STD ERR	754.905	MEDIAN	8201.658
MODE	.344	STD DEV	34684.580	VARIANCE	1203020107
KURTOSIS	13.492	S E KURT	.106	SKEWNESS	2.884
S E SKEW	.053	RANGE	383960.594	MINIMUM	.344
MAXIMUM	383960.938	SUM	50076632.3		

6.33 The nature of sliver polygons.

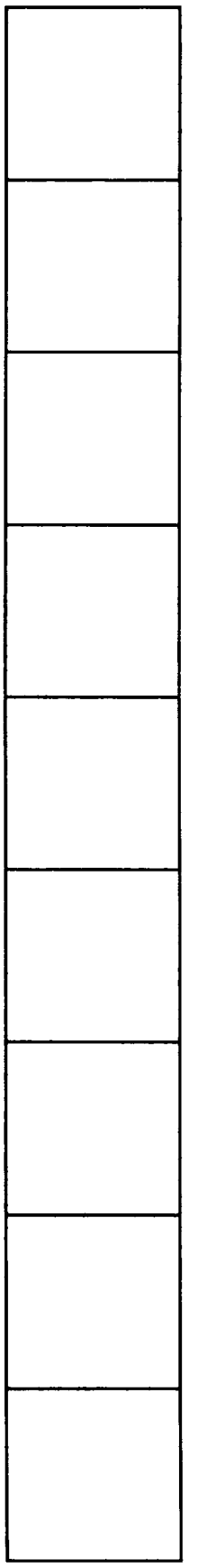


The classic shape of sliver polyons will be long and thin

6.34 Area as an unreliable parameter for sliver polygon removal.



Polygon a



Polygon b

The Areal extent of both the square and the long thin strip are the same; 9 units. Thus, using area alone as a potential indicator of sliver polygons may remove polygons which are intentionally small, and which do not conform to the classic sliver shape.

6.35 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of PERIMETER: 1,000 units

6.36 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of PERIMETER 2,000 unit

6.37 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of PERIMETER 2,500 unit

6.38 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA 5,000 units

6.39 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA: 8,000 units

6.40 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA: 150,000 units

6.41 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA 50,000 units

6.42 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER: 10 units

6.43 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER: 12 units

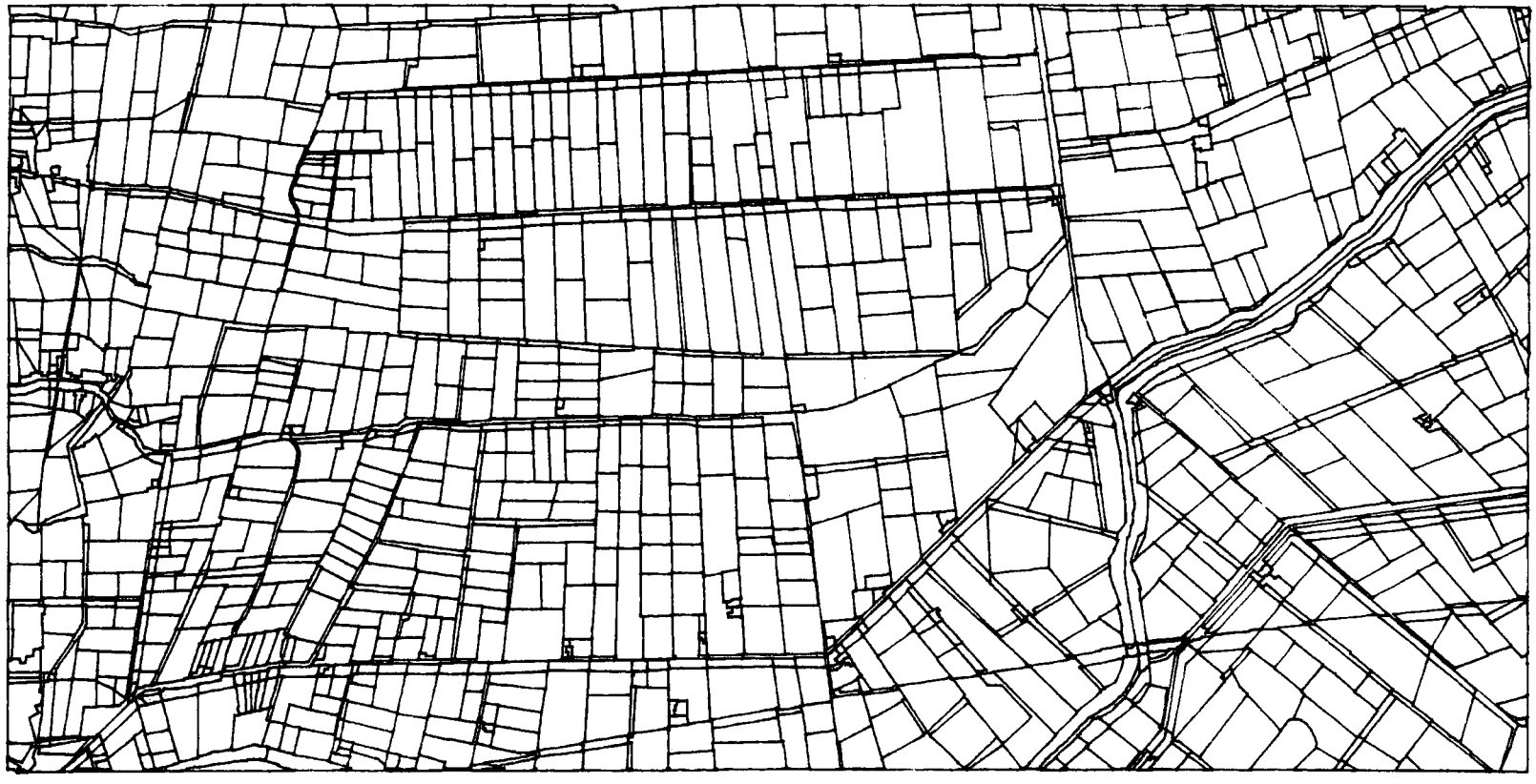
6.44 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER: 15 units

6.45 An overlay of the 1:50,000 coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER: 20 units

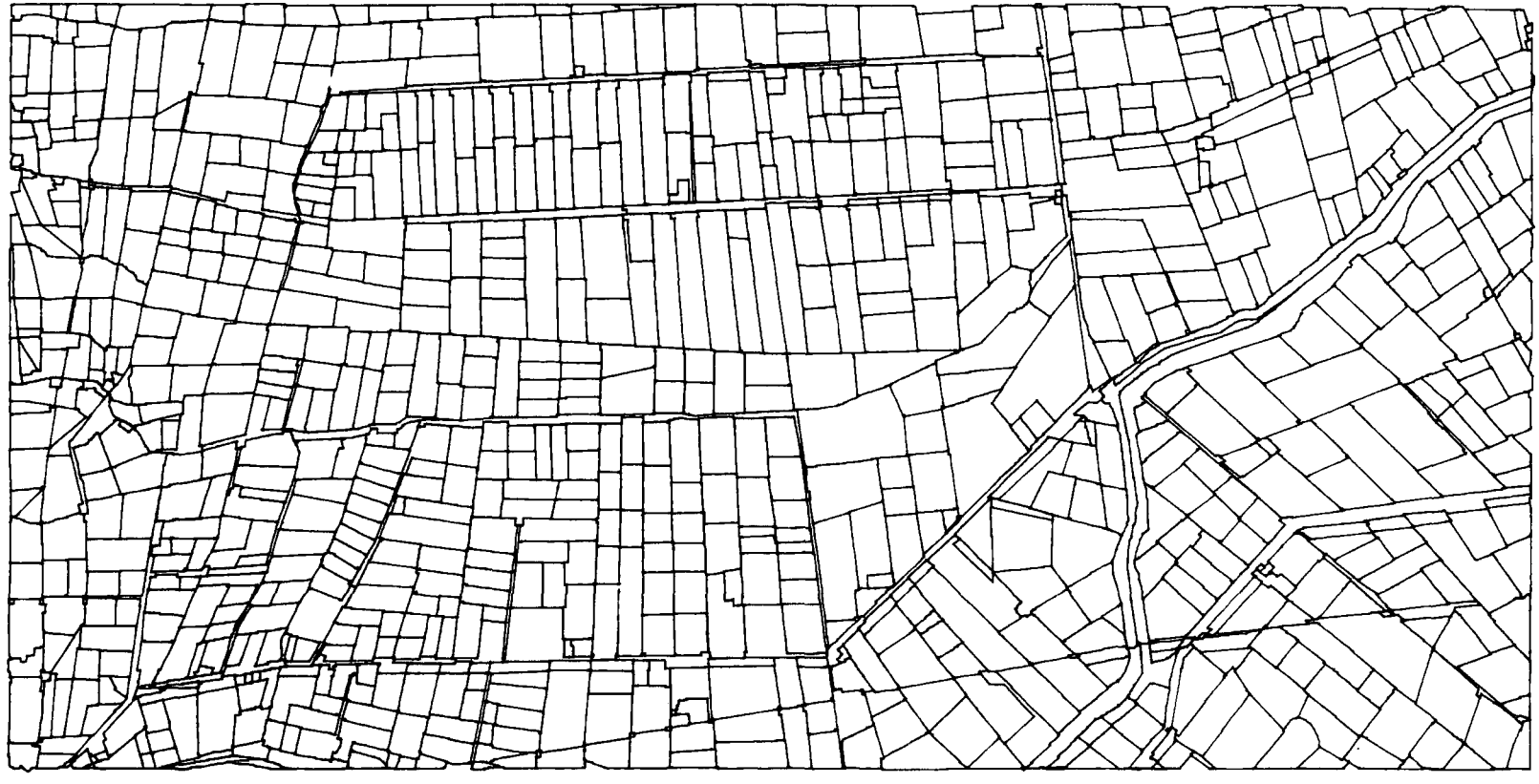
6.46 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER: 25 units

6.47 An overlay of the 1:50,000a coverage and the 1:50,000b coverage following an ELIMINATE of AREA/PERIMETER: 30 units

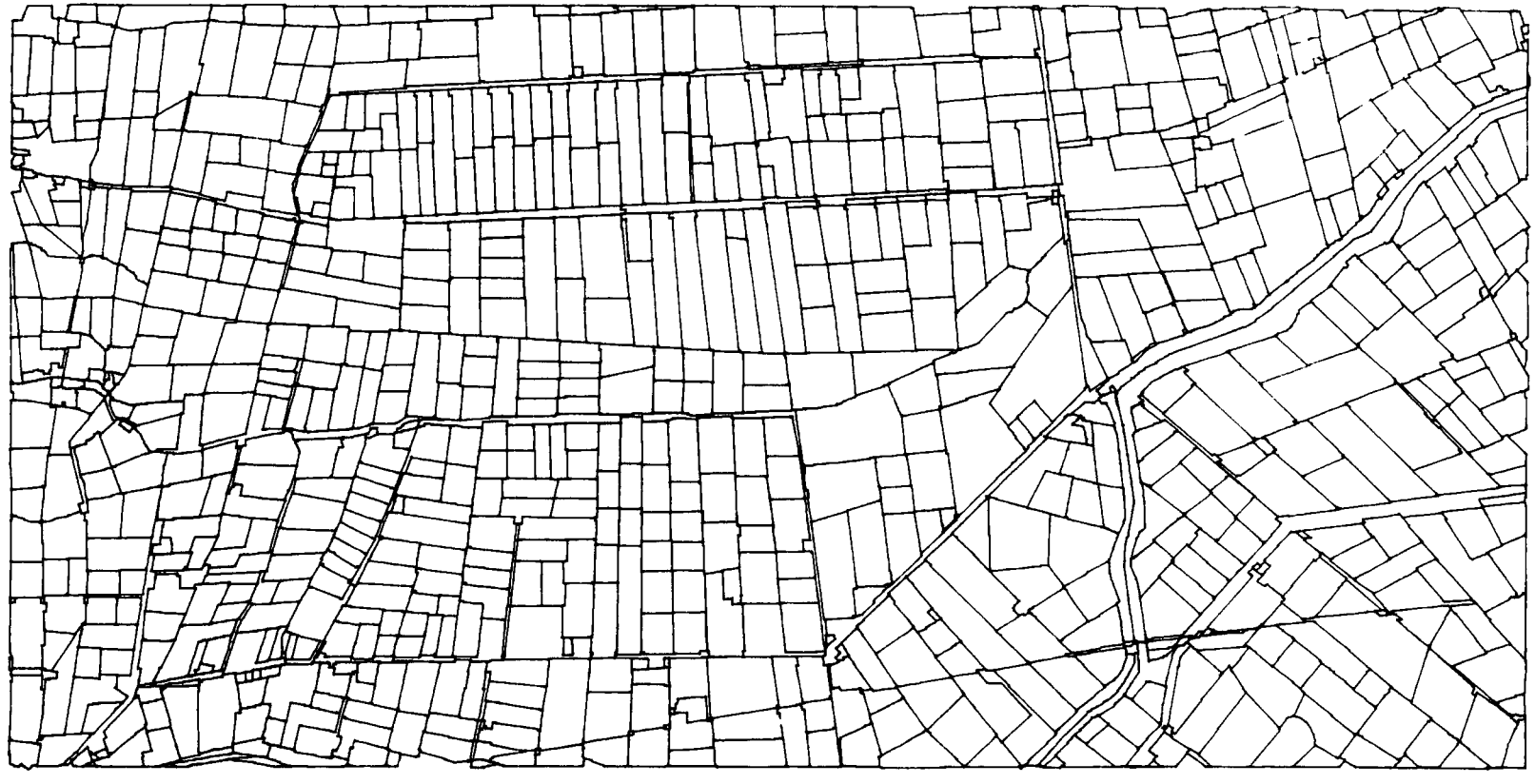
6.48 An overlay of the 1:50,000a coverage and the 1:50,0000b coverage following an ELIMINATE of AREA/PERIMETER: 50 units



6.49 An overlay of the 1:50,000b coverage, the 1:25,000 coverage and the 1:10,560 coverage following an ELIMINATE of AREA/PERIMETER: 10 units



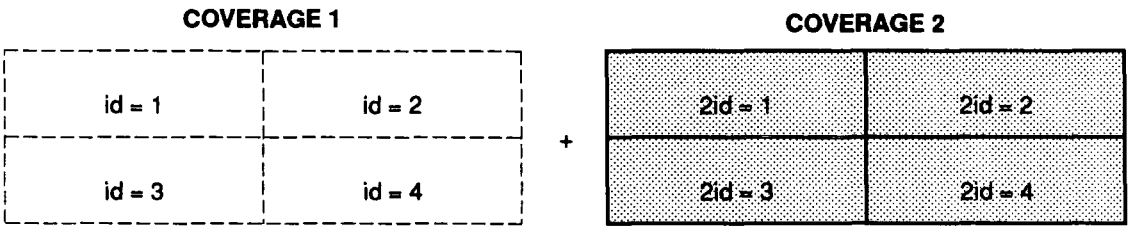
6.50 An overlay of the 1:50,000b coverage, the 1:25,000 coverage and the 1:10,560b coverage following an ELIMINATE of AREA/PERIMETER: 20 units



6.51 An overlay of the 1:50,000b coverage, the 1:25,000 coverage and the 1:10,560b coverage following an ELIMINATE of AREA/PERIMETER: 30 units

Figures for Chapter 7

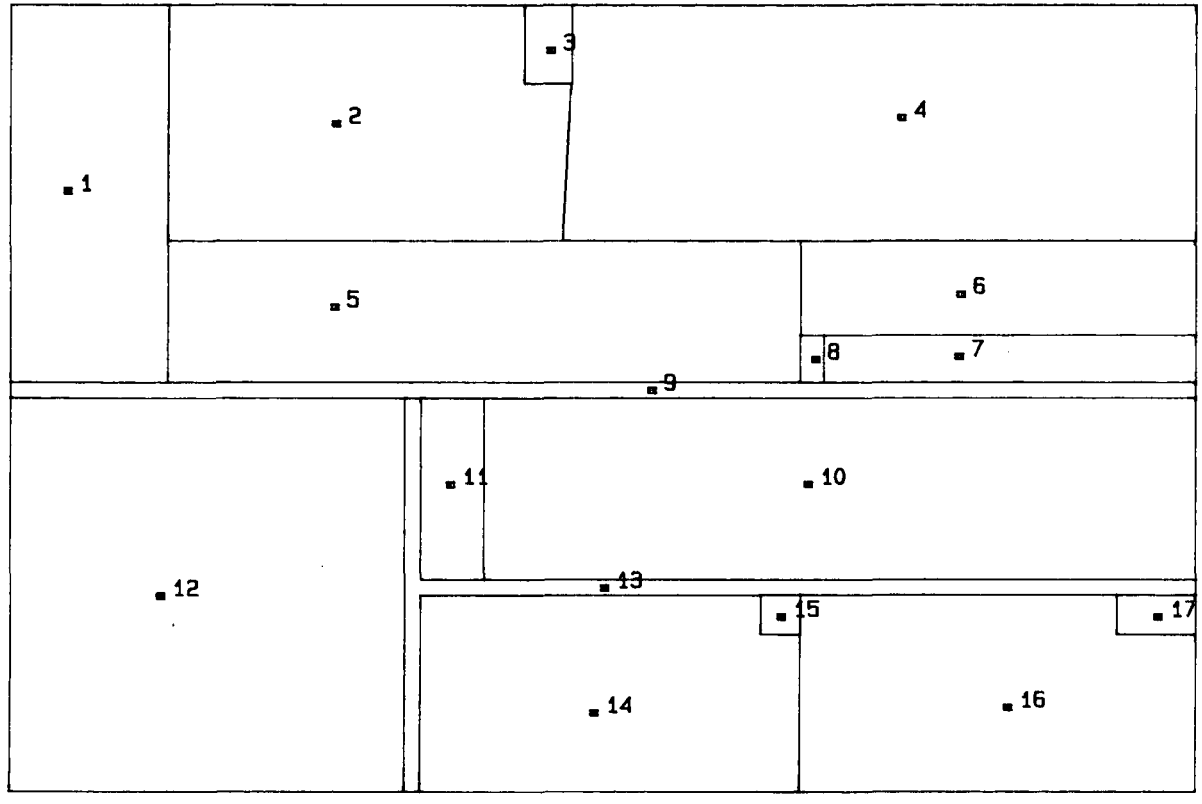
7.1 Using polygon identifiers as a basis for overlay error detection



OVERLAY OF COVERAGE 1 AND 2

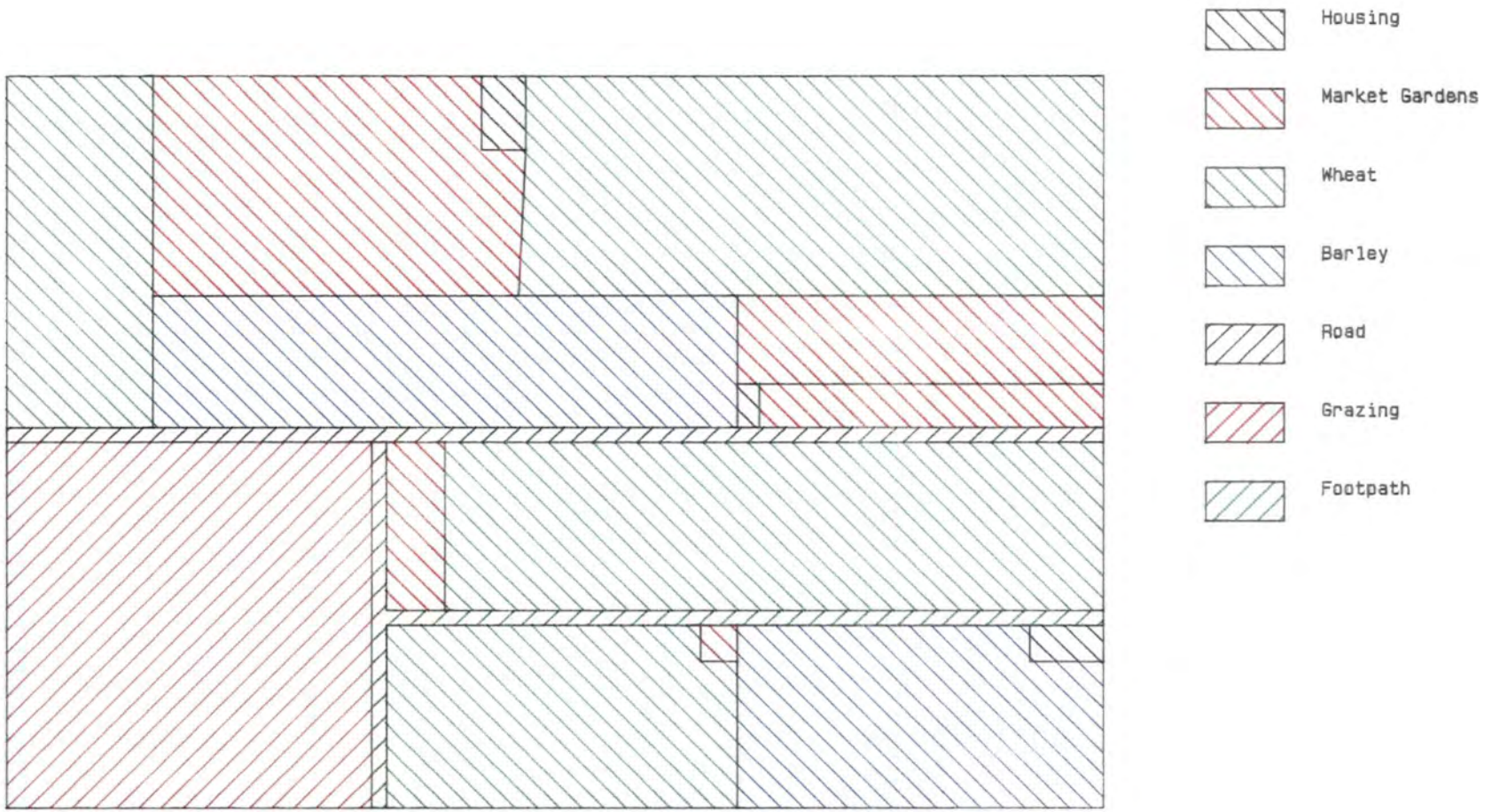
id = 1 2id = ϕ	id = 2 2id = ϕ
id = 1 2id = 1	id = 2 2id = 2
id = 3 2id = 1	id = 4 2id = 2
id = 3 2id = 3	id = 4 2id = 4
id = ϕ 2id = 3	id = ϕ 2id = 4

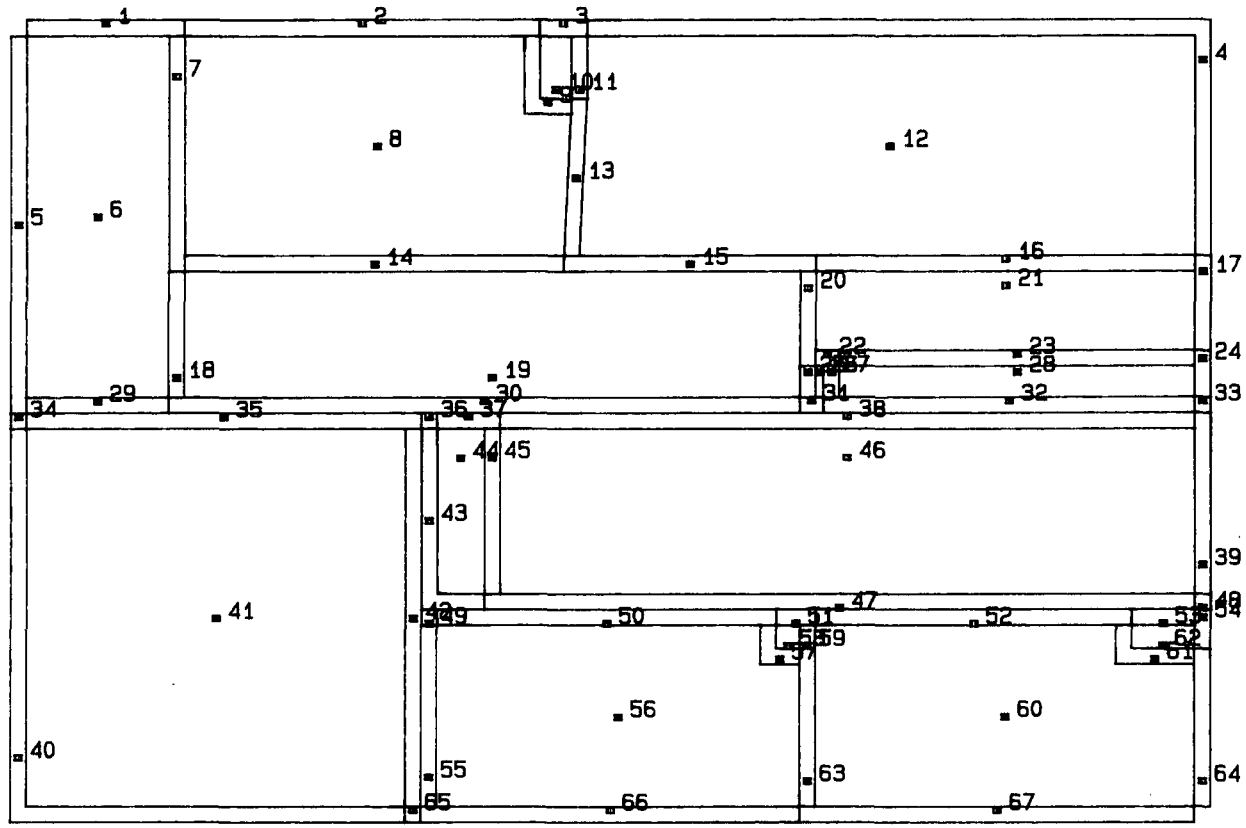
Coverage 1 and 2 have the same geometrical boundaries and these should be identical. Due to problems of overlay, slivers arise but these can then be removed by forcing the software to join units carrying the same identifiers



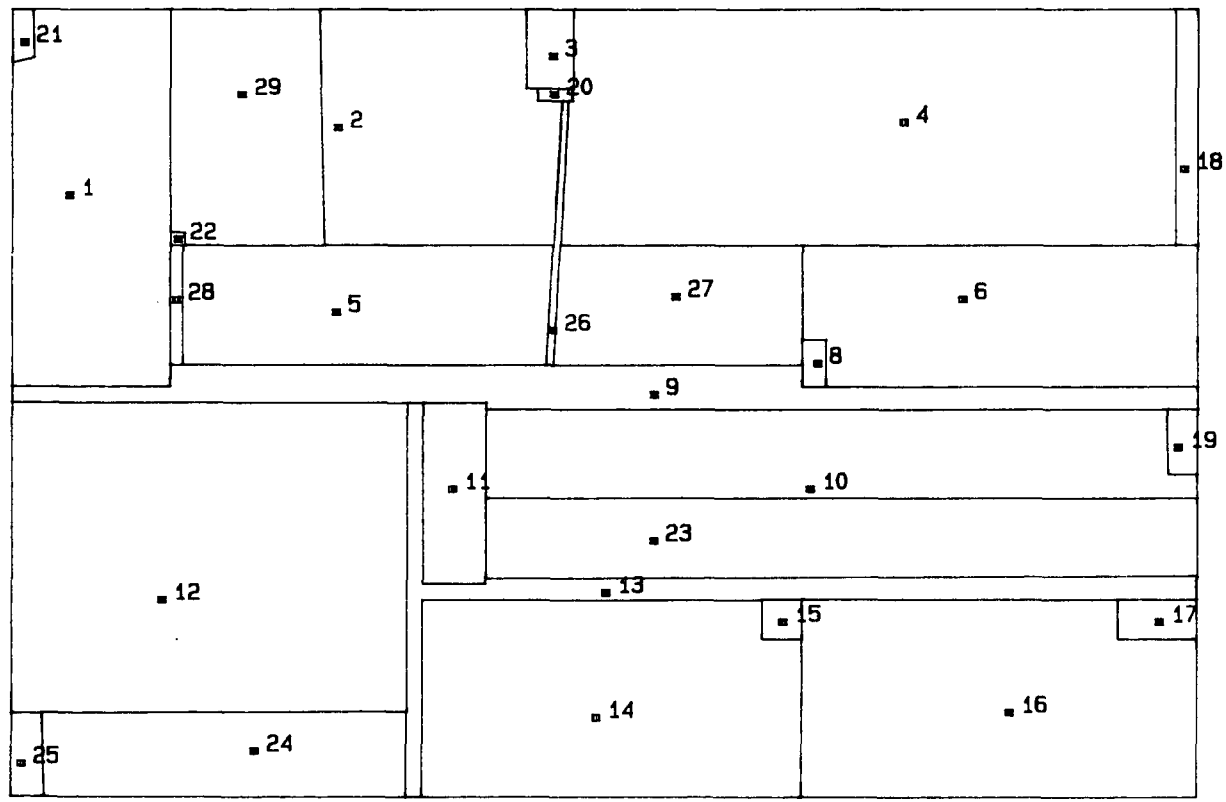
7.2 A test coverage consisting of small zones and long thin zones

7.3 The test coverage showing assigned land use values

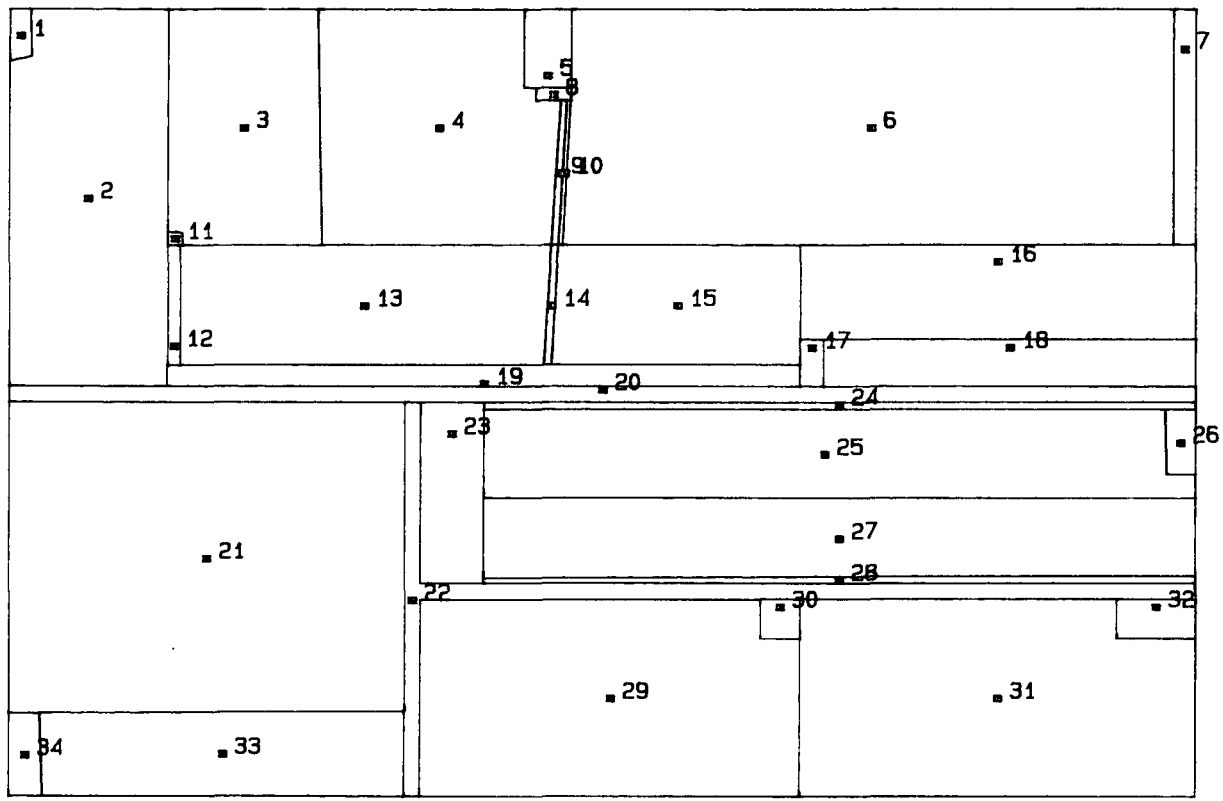




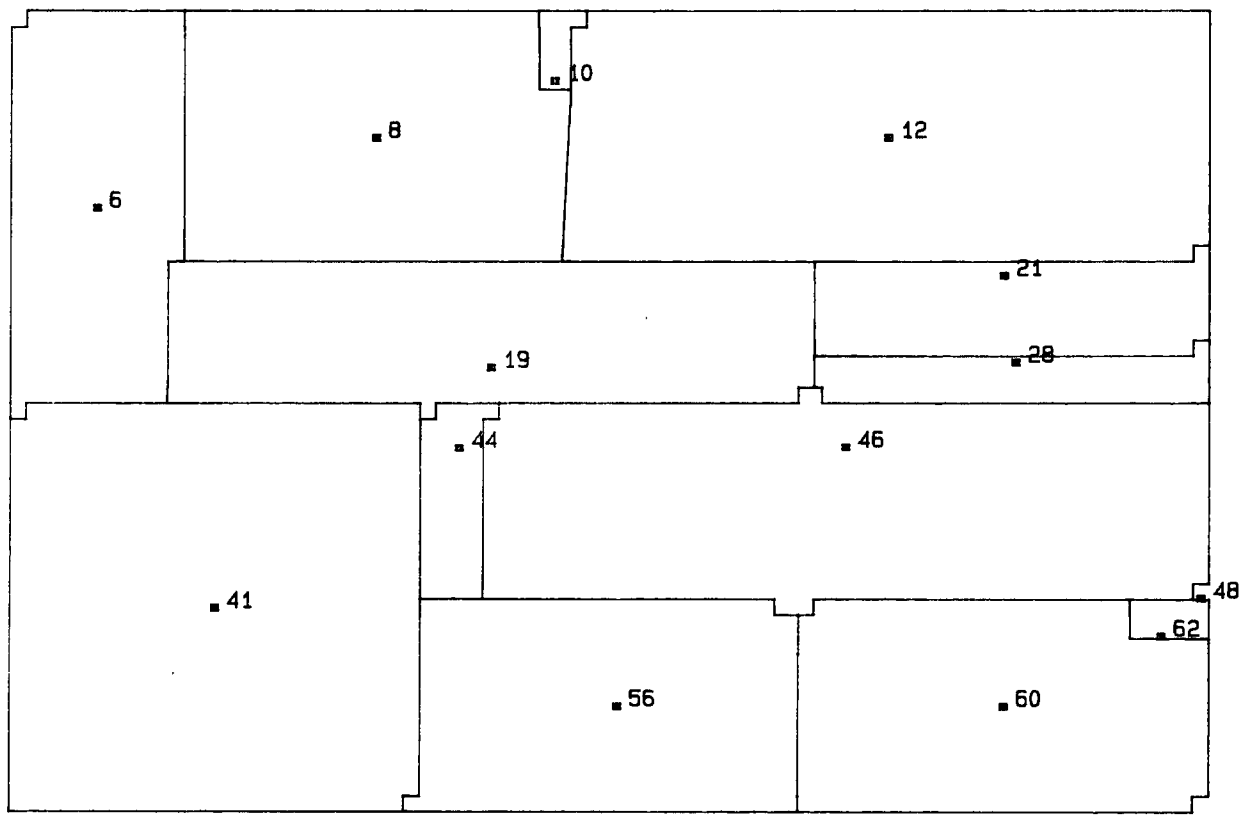
7.4 Overlay of the original coverage and the transformed coverage each consisting of the same features



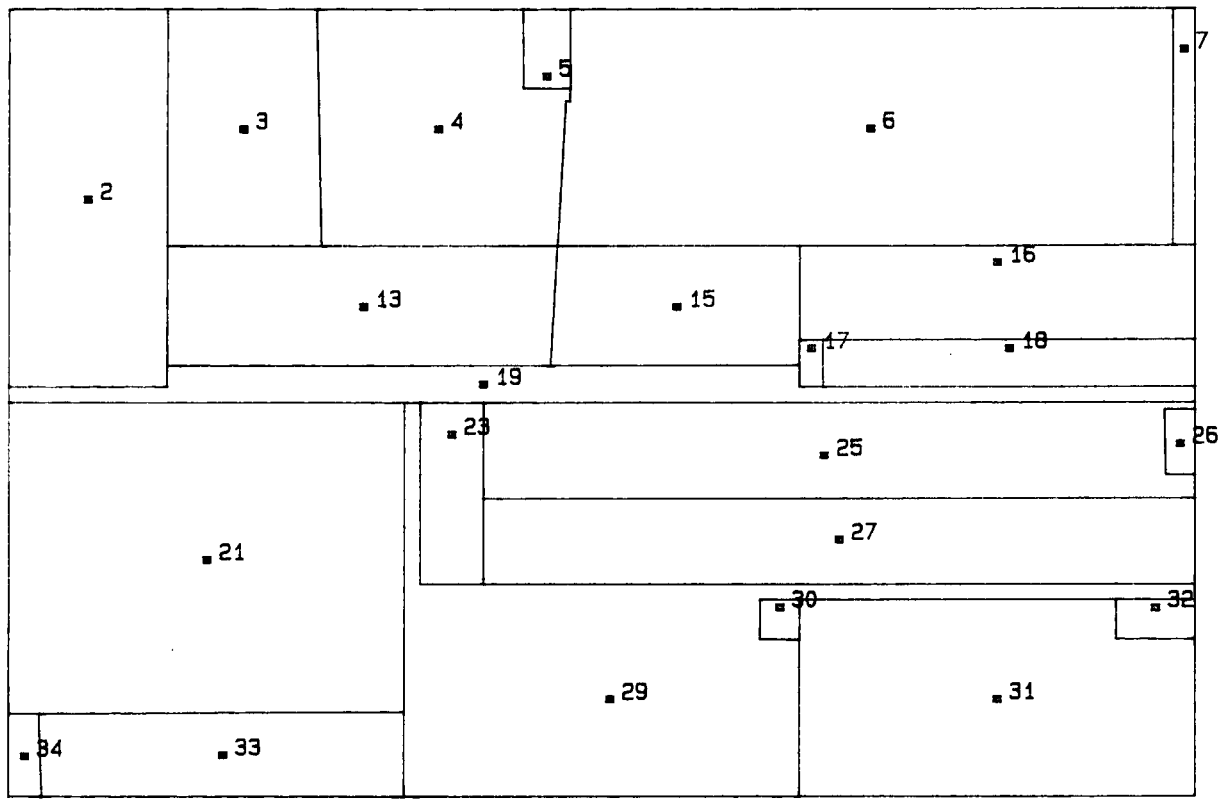
7.5 Coverage of change



7.6 Overlay of the original coverage and the coverage of change

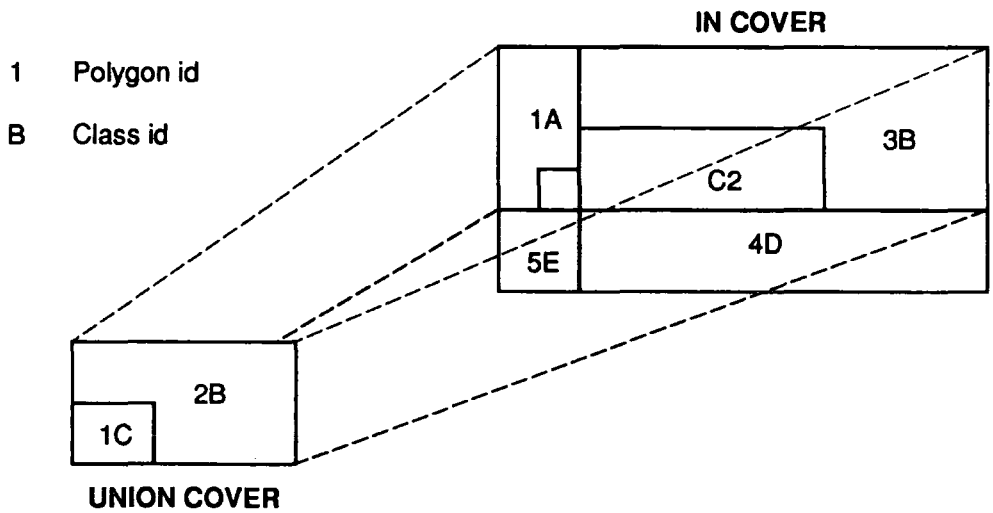


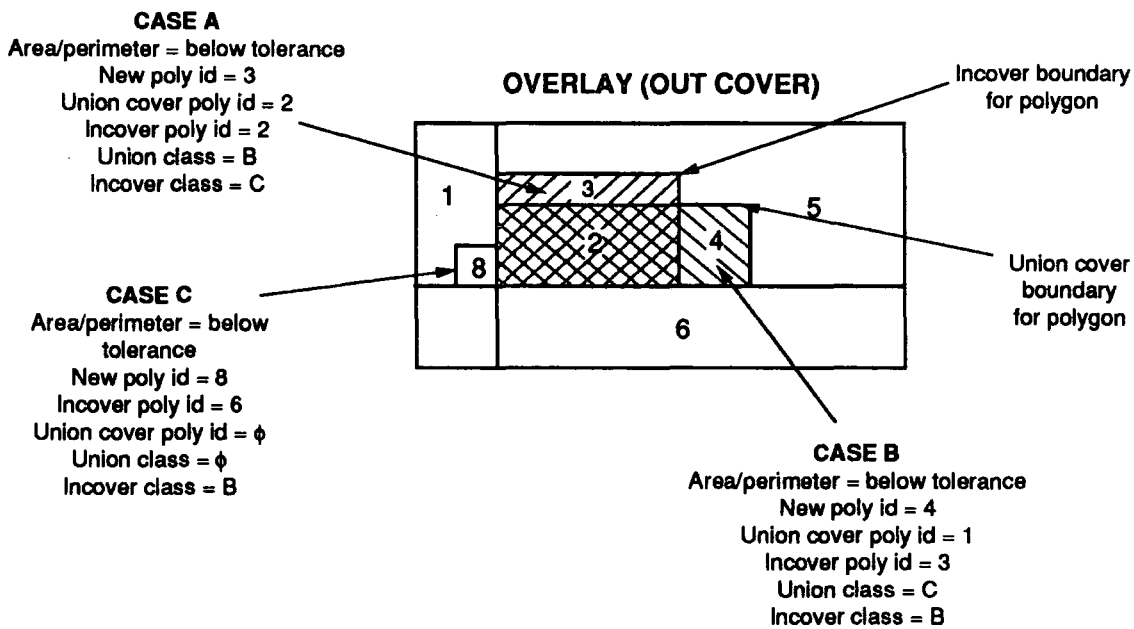
7.7 ELIMINATION of errors by forcing coverages to match on the basis of equal identifiers



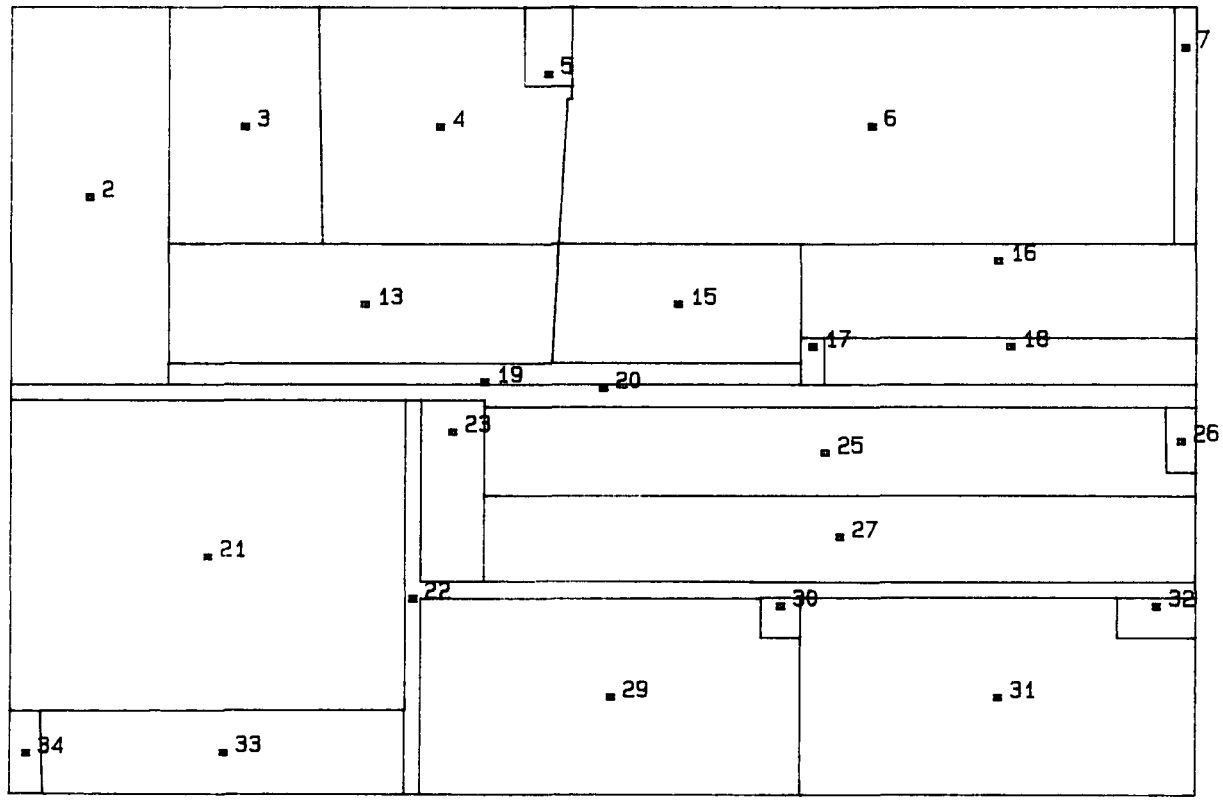
7.8 ELIMINATION on the basis of AREA/PERIMETER less than 0.1 units

7.9 Determining allowable movement on the basis of a priori knowledge of the data

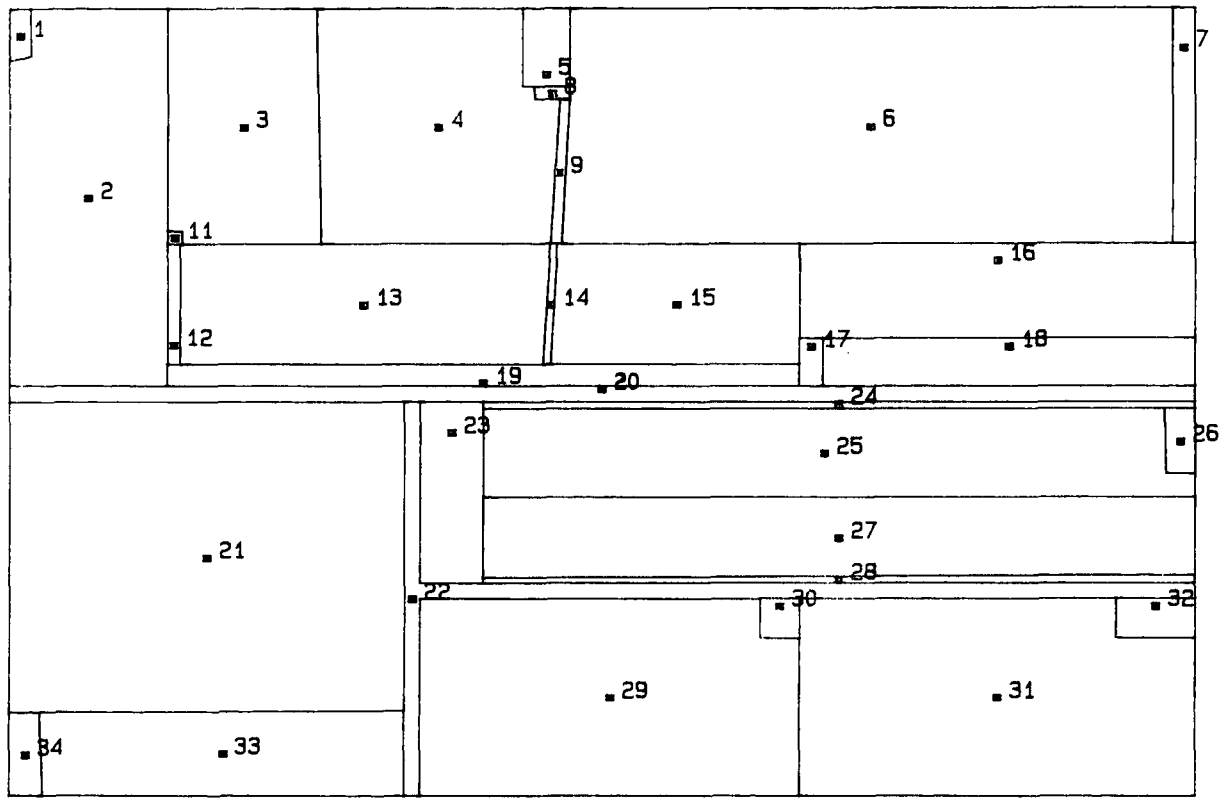




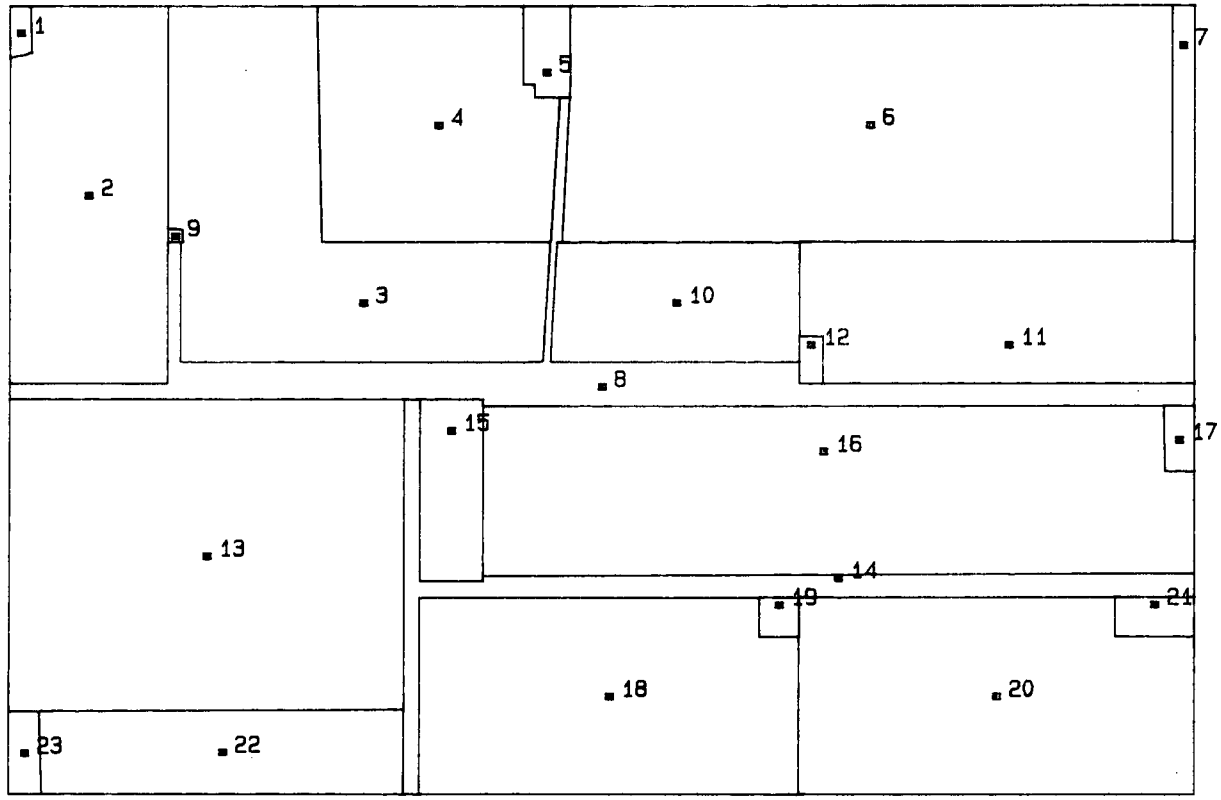
Polygon movement is only allowed between classes in certain predefined combinations. For example if incov class = B and union class = C; the old class will be retained **BUT** if incov class = C and the union class = B; then the new class is retained



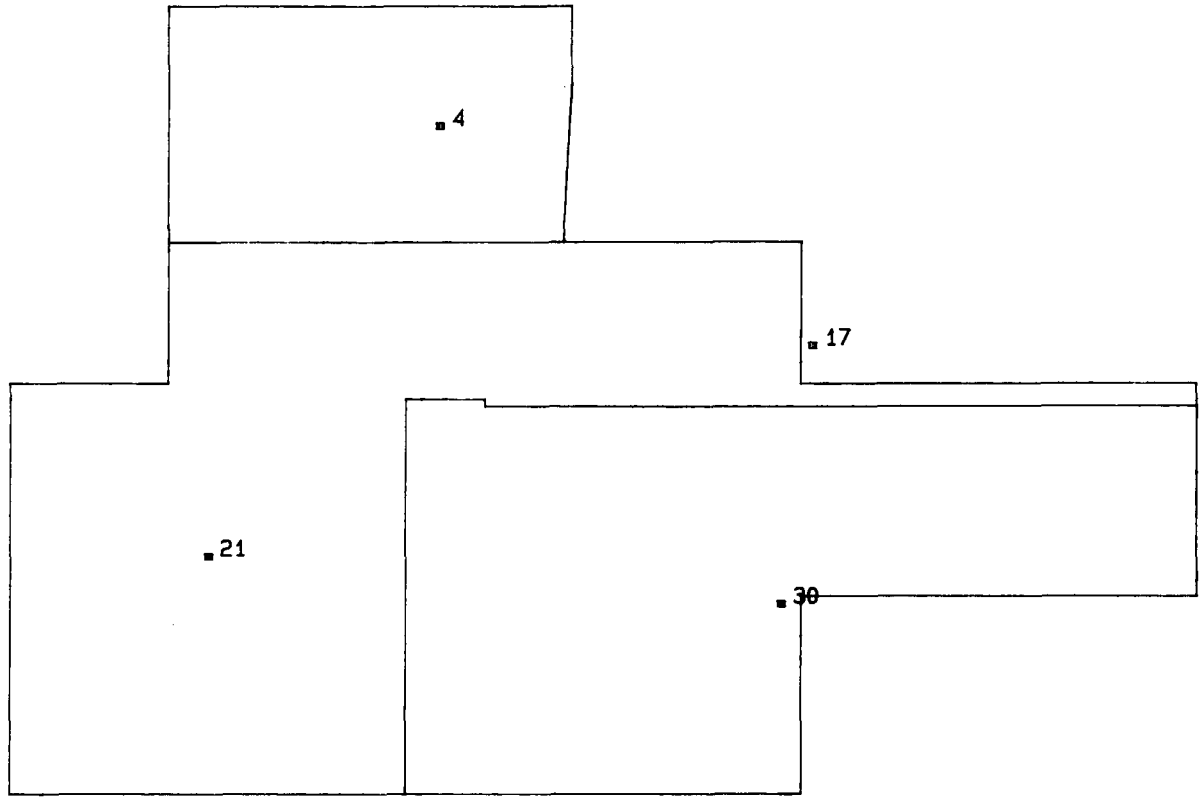
7.10 Using CLASS as the basis for ELIMINATION



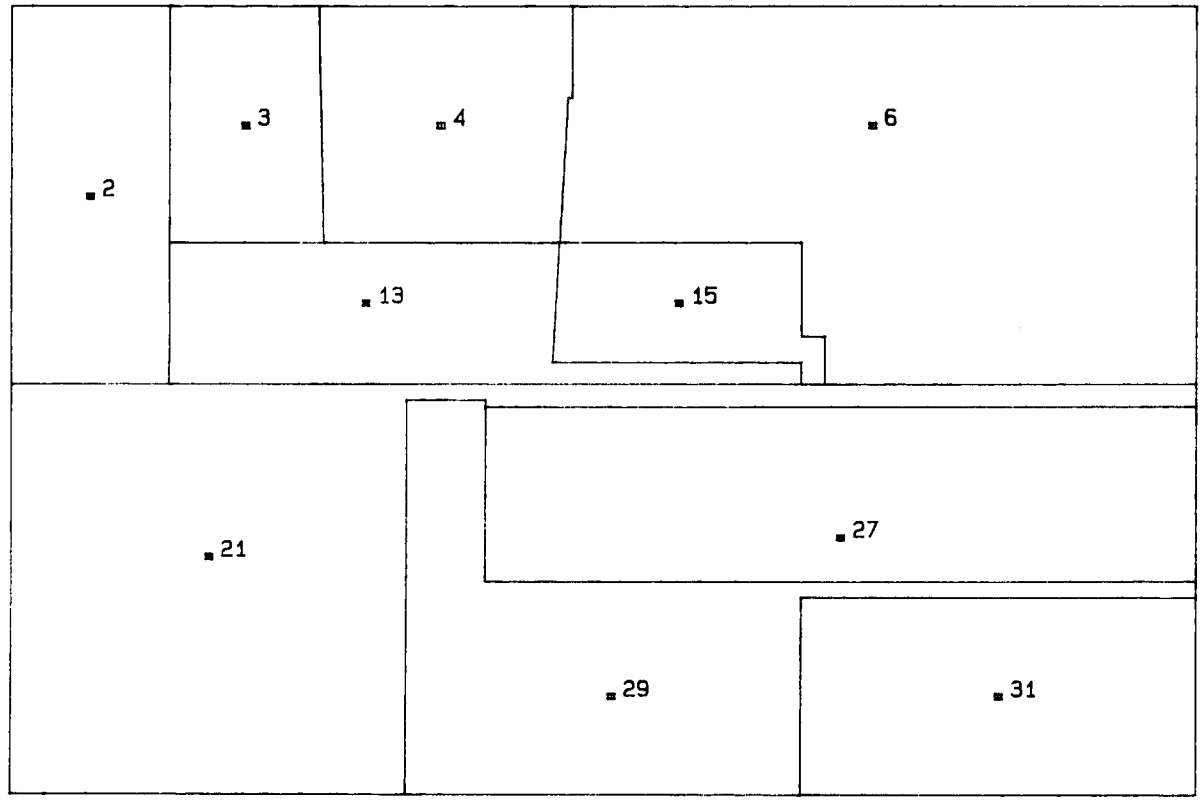
7.11 Using either CLASS or CLASS2 as the basis for ELIMINATION



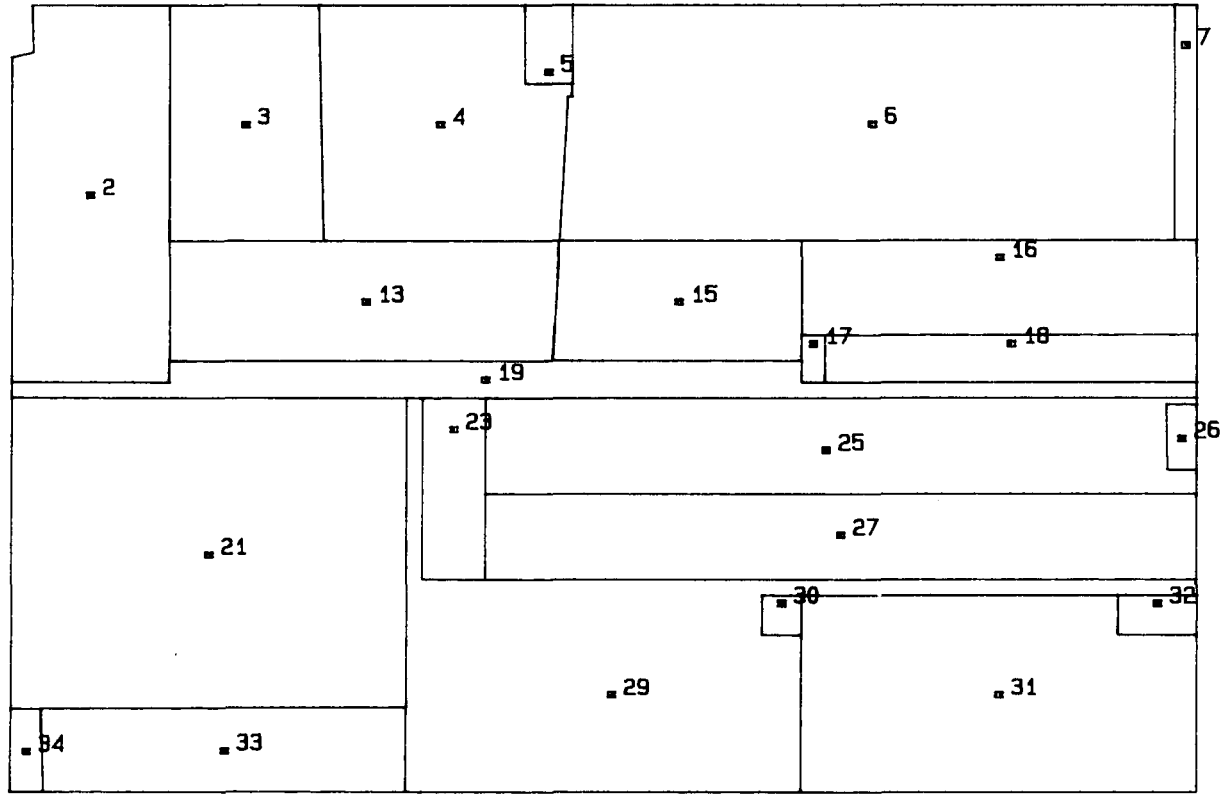
7.12 Dissolving on the CLASS2 attribute following the use of CLASS or CLASS2 as the basis for ELIMINATION



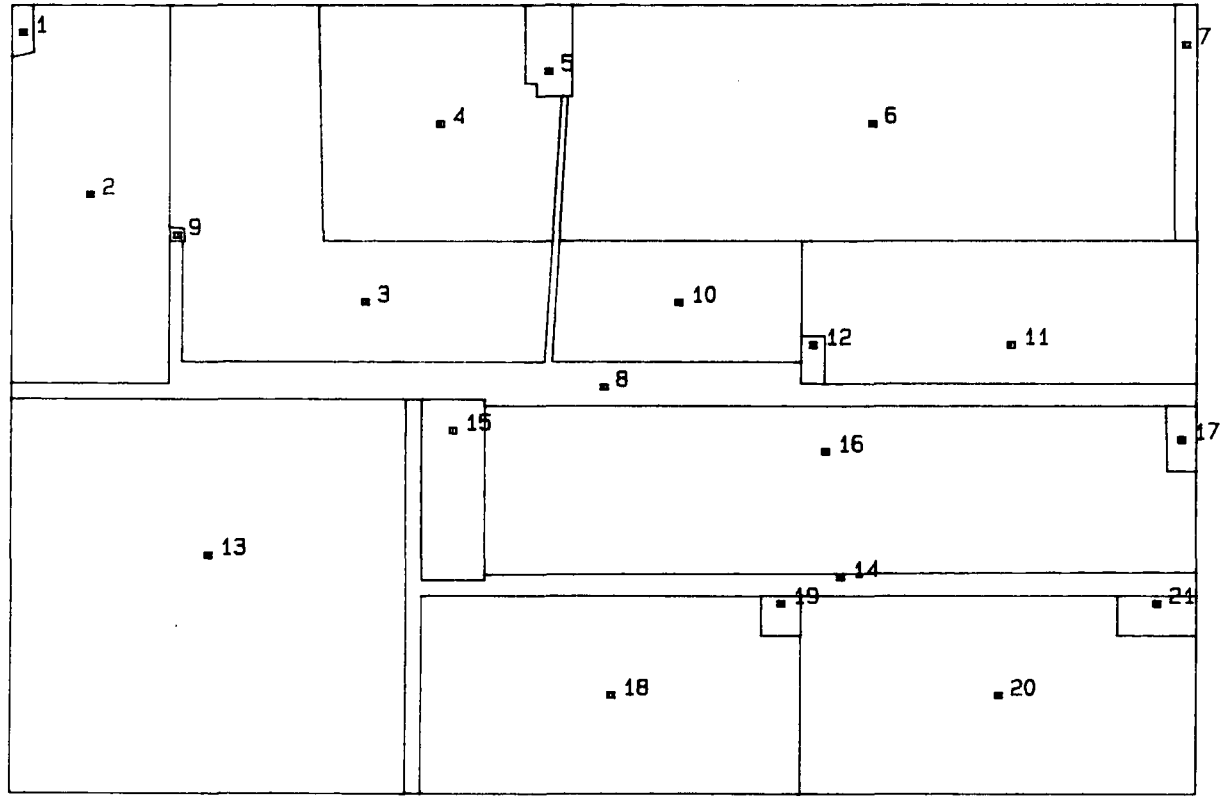
7.13 ELIMINATION on the basis of AREA/PERIMETER being less than 10 units



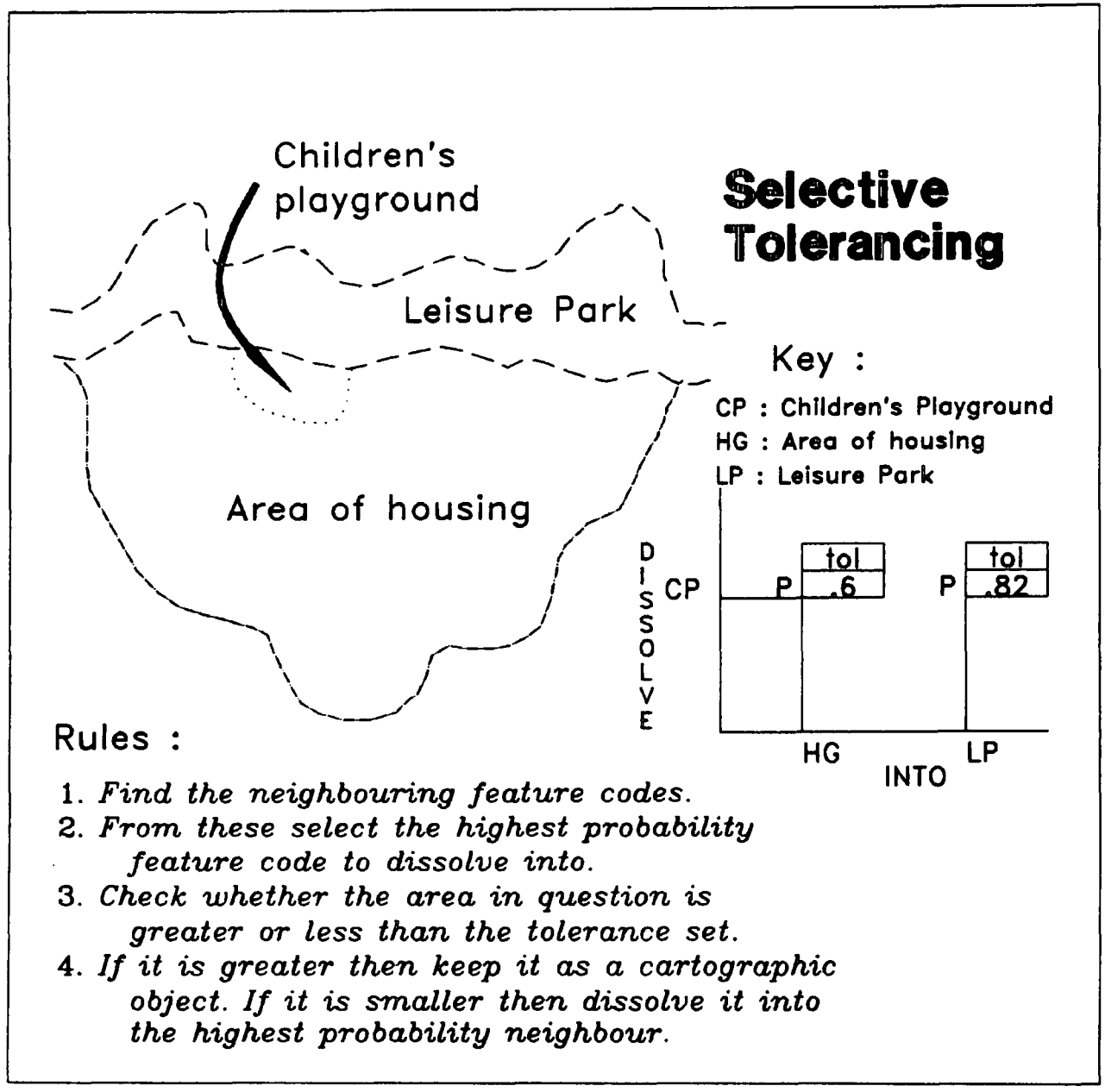
7.14 ELIMINATION on the basis of AREA/PERIMETER being less than 0.5 units



7.15 ELIMINATION on the basis of PERIMETER/AREA being greater than 10

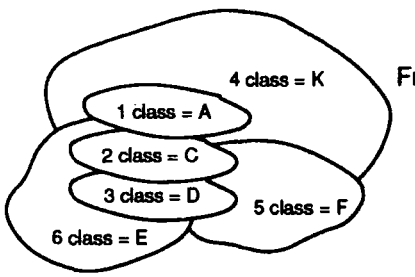


7.16 Coverage dissolved on the basis of database rules without the use of the ELIMINATE option



7.17 A methodology for selective neighbourhood matching for sliver polygons

THE ORIGINAL COVERAGE

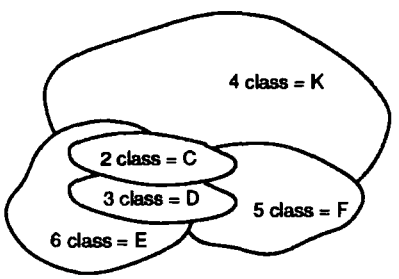


RULES

From	to		Prob
A	K	=	60%
A	C	=	50%
A	D	=	40%
A	E	=	30%
A	F	=	20%

Sliver 1 is dealt with first; it falls below the tolerance and is merged with sliver 4 on the basis of the rules. The neighbourhood relationships between polygon 1 and other polygons are removed and appended onto polygon 4

THE RESULT OF POLYGON 1 BEING MERGED WITH POLYGON 4

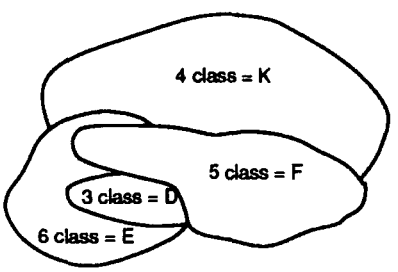


From	to		Prob
C	K	=	40%
C	A	=	70%
C	F	=	50%
C	E	=	20%
C	D	=	10%

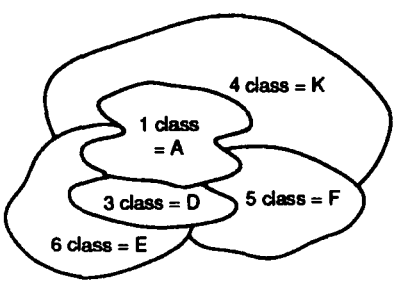
Polygon 2 = class C

Sliver 2 is dealt with second; it falls below the tolerance and is merged with polygon 5 on the basis of the rules

THE RESULT OF POLYGON 2 BEING MERGED WITH POLYGON 5



THE RESULT OF POLYGON 2 BEING MERGED WITH POLYGON 1

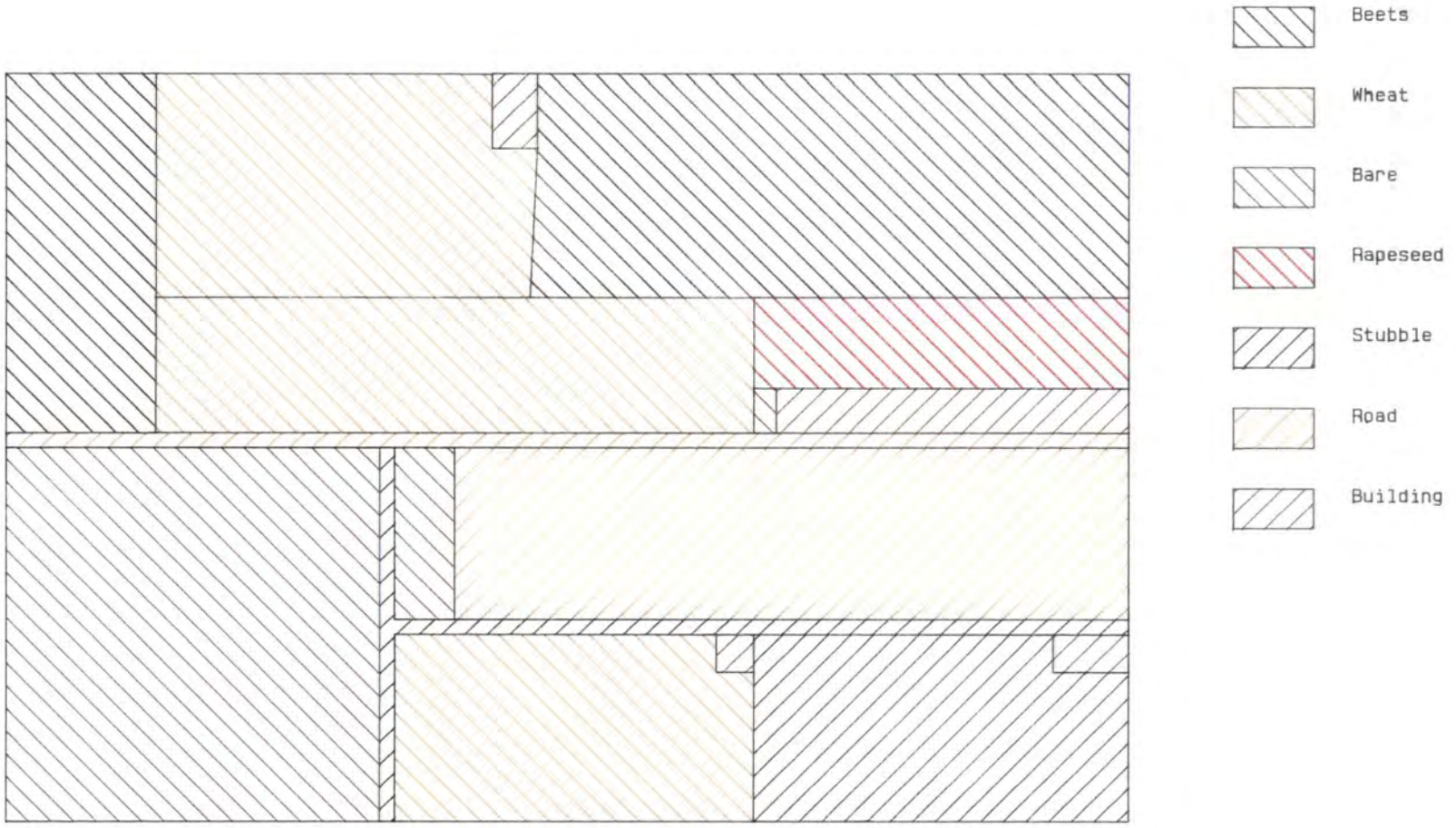


From	to		Prob
C	K	=	40%
C	A	=	70%
C	F	=	50%
C	E	=	20%
C	D	=	10%

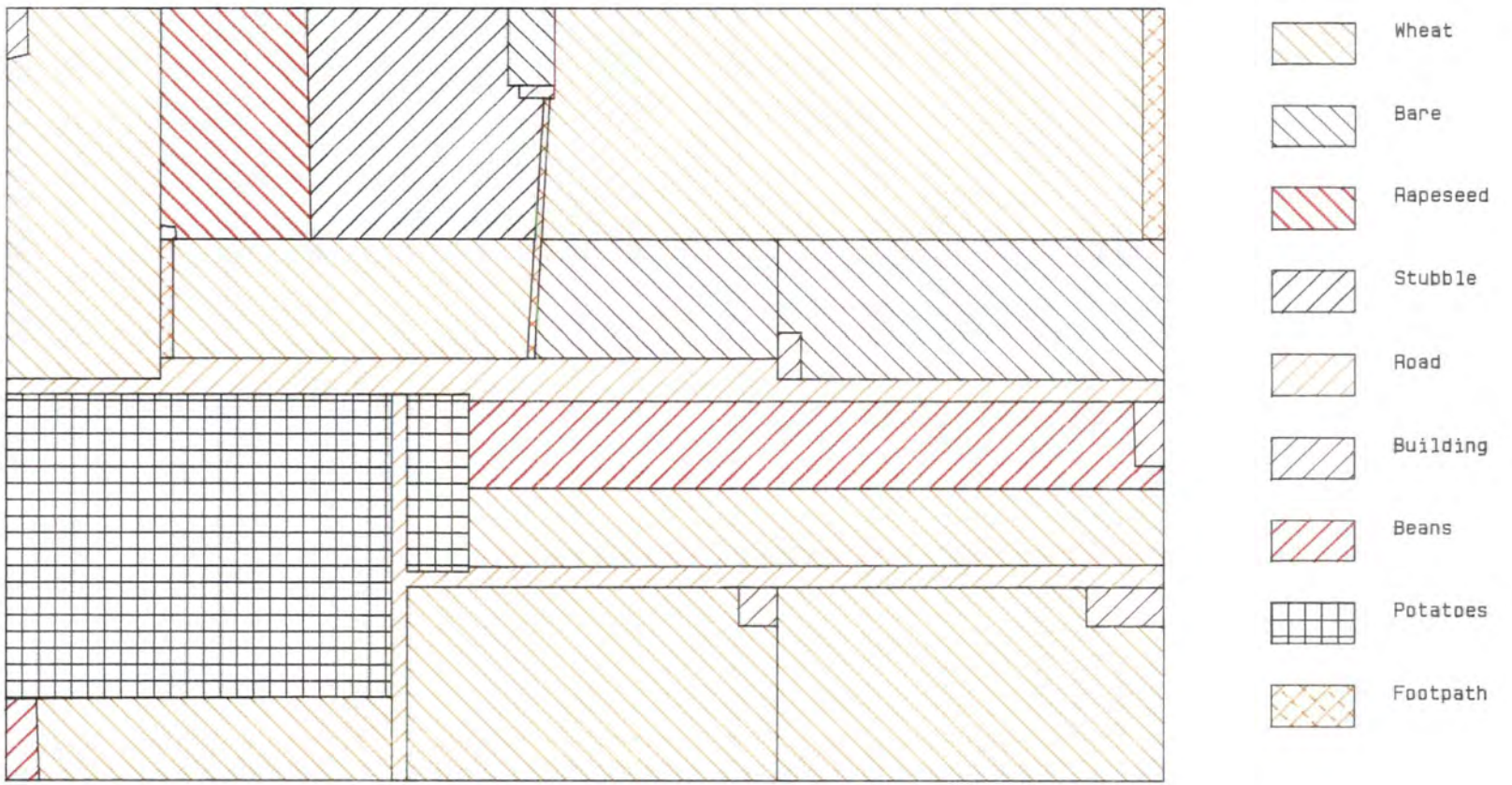
Polygon 2 = class C

If polygon 2 had been dealt with first it would have merged with polygon 1 on the basis of the rules. Polygon 1 would not have fallen below the tolerance and a different land cover pattern would have ensured

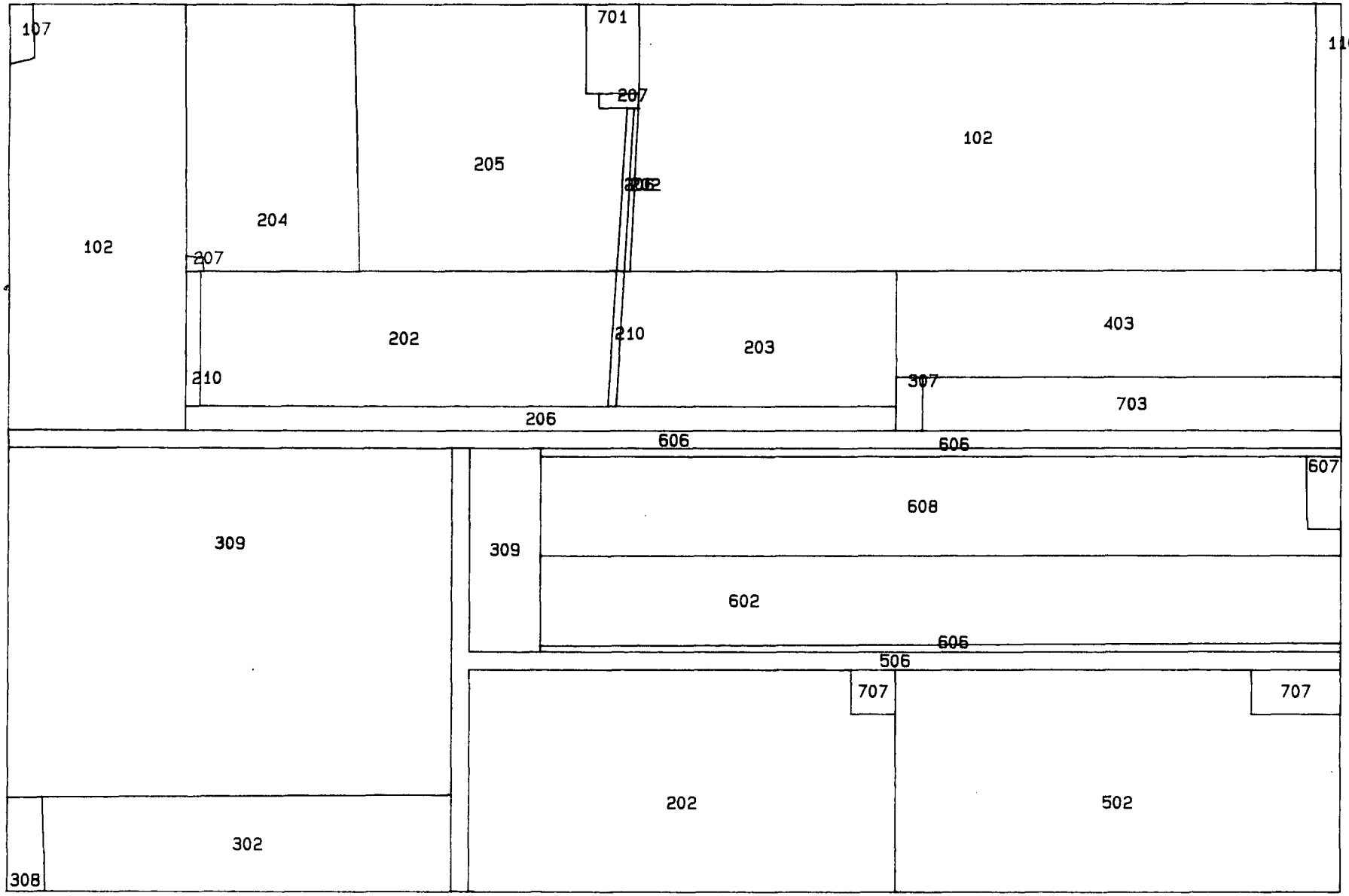
7.19 Land use categories assigned to the first coverage

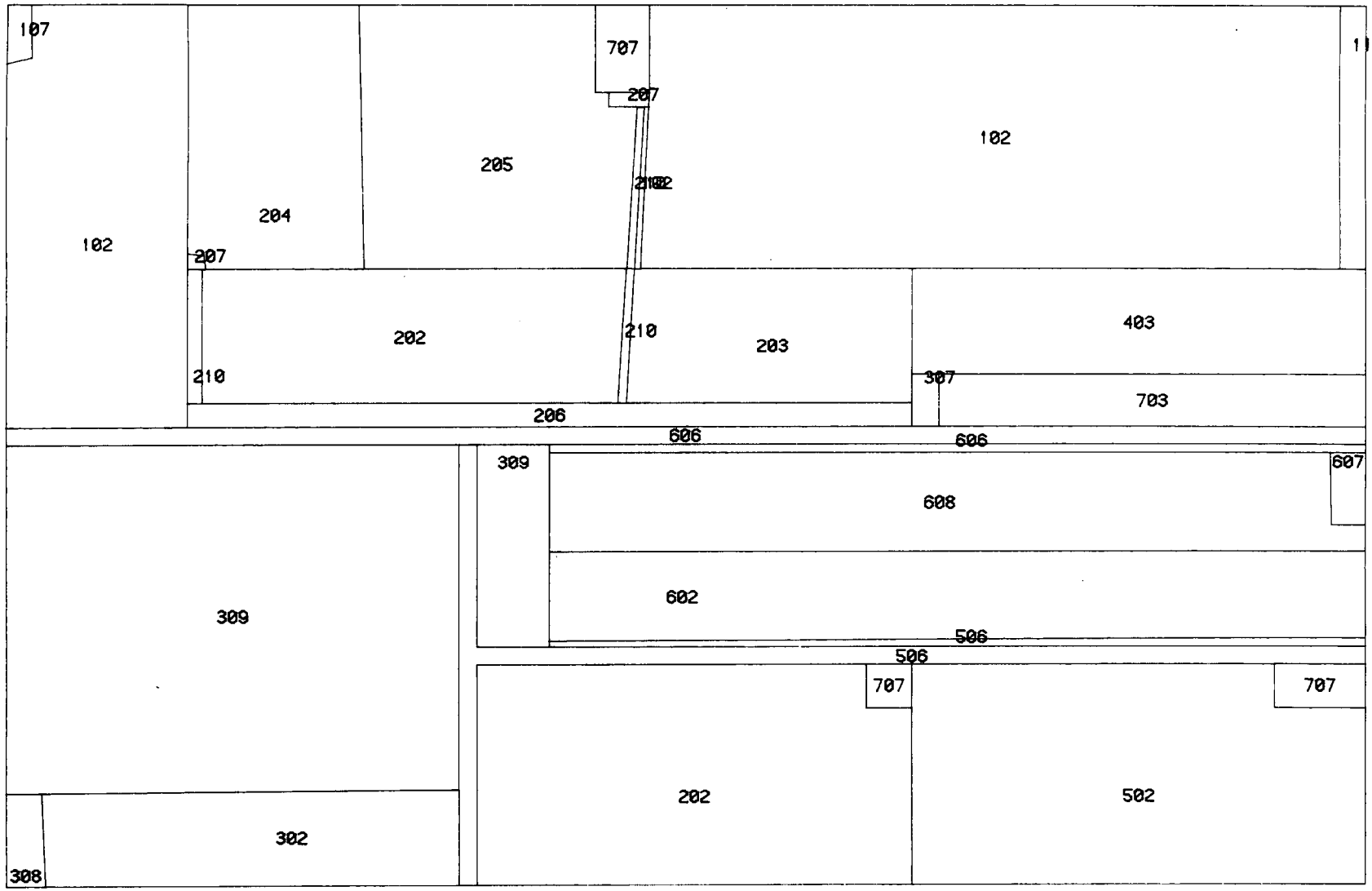


7.20 Land use categories assigned to the overlay of change

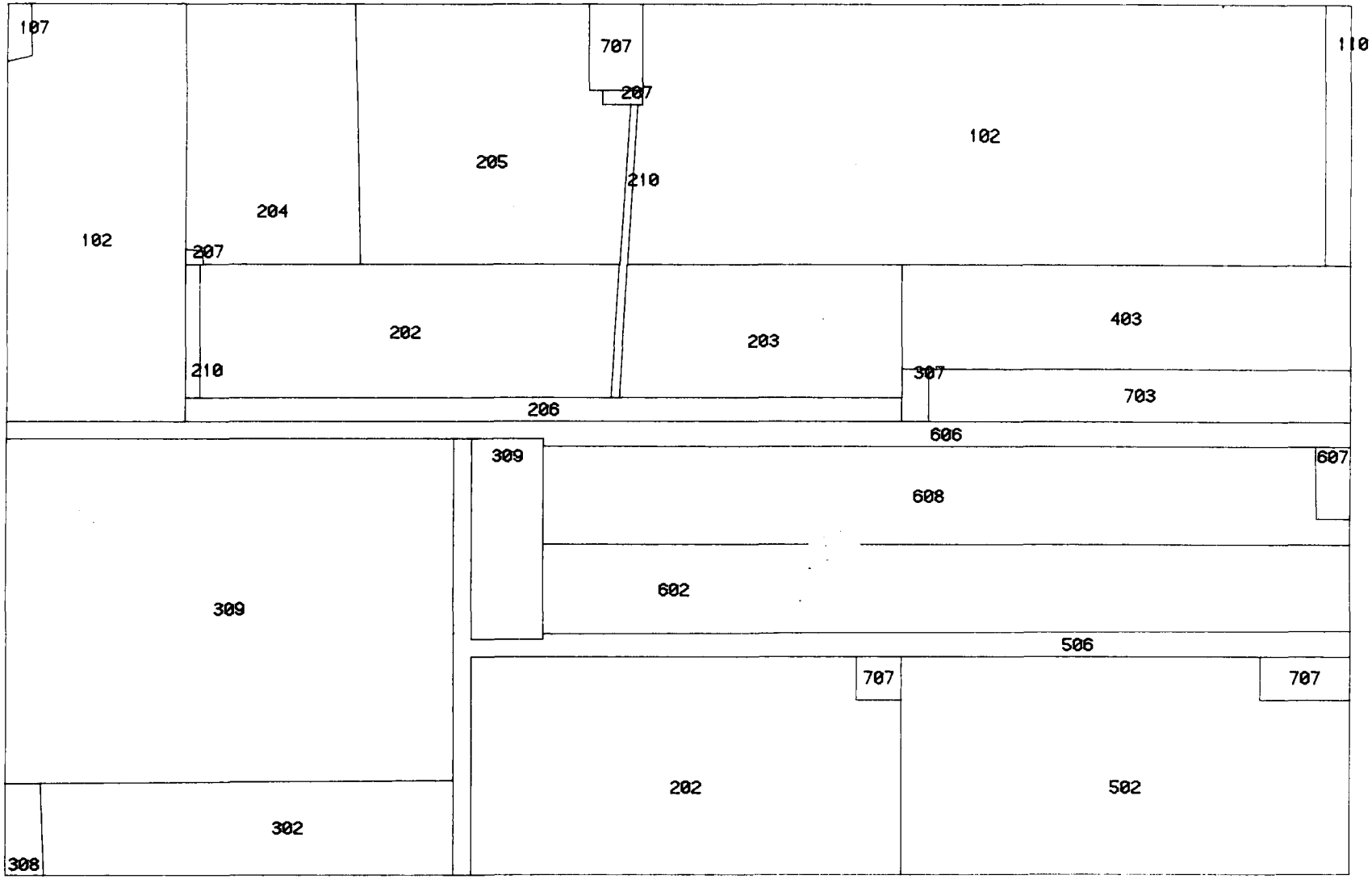


7.21 The overlaid coverages showing their histories





7.22 The overlaid coverages showing changed histories as defined by the matrix of probable histories



7.23 The overlaid coverages following a DISSOLVE operation on the new polygon history values