A Removal Set Based Approach to Reliability Assessment in Urban Water Supply Distribution Networks

Paul Jacobs

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment

of the Requirements for the Degree of Doctor of Philosophy

Department of Civil Engineering

The University of Manitoba

Winnipeg, Manitoba

ISBN 0-315-81793-3

Name ____

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Civil Engineering | 0543 | U·M·I

SUBJECT TERM | SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences
Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984

Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

Nom ____________________

Dissertation Abstracts International est organisé en catégories de sujets. Veuillez s.v.p. choisir le sujet qui décrit le mieux votre thèse et inscrivez le code numérique approprié dans l'espace réservé ci-dessous.

SUJET ____________________ CODE DE SUJET ☐☐☐☐ U·M·I

Catégories par sujets

HUMANITÉS ET SCIENCES SOCIALES

COMMUNICATIONS ET LES ARTS
Architecture 0729
Beaux-arts 0357
Bibliothéconomie 0399
Cinéma 0900
Communication verbale 0459
Communications 0708
Danse 0378
Histoire de l'art 0377
Journalisme 0391
Musique 0413
Sciences de l'information 0723
Théâtre 0465

ÉDUCATION
Généralités 515
Administration 0514
Art 0273
Collèges communautaires 0275
Commerce 0688
Économie domestique 0278
Éducation permanente 0516
Éducation préscolaire 0518
Éducation sanitaire 0680
Enseignement agricole 0517
Enseignement bilingue et multiculturel 0282
Enseignement industriel 0521
Enseignement primaire. 0524
Enseignement professionnel 0747
Enseignement religieux 0527
Enseignement secondaire 0533
Enseignement spécial 0529
Enseignement supérieur 0745
Évaluation 0288
Finances 0277
Formation des enseignants 0530
Histoire de l'éducation 0520
Langues et littérature 0279
Lecture 0535
Mathématiques 0280
Musique 0522
Orientation et consultation 0519
Philosophie de l'éducation 0998
Physique 0523
Programmes d'études et enseignement 0727
Psychologie 0525
Sciences 0714
Sciences sociales 0534
Sociologie de l'éducation 0340
Technologie 0710

LANGUE, LITTÉRATURE ET LINGUISTIQUE
Langues
- Généralités 0679
- Anciennes 0289
- Linguistique 0290
- Modernes 0291

Littérature
- Généralités 0401
- Anciennes 0294
- Comparée 0295
- Mediévale 0297
- Moderne 0298
- Africaine 0316
- Américaine 0591
- Anglaise 0593
- Asiatique 0305
- Canadienne (Anglaise) 0352
- Canadienne (Française) 0355
- Germanique 0311
- Latino-américaine 0312
- Moyen-orientale 0315
- Romane 0313
- Slave et est-européenne 0314

PHILOSOPHIE, RELIGION ET THÉOLOGIE
Philosophie 0422
Religion
- Généralités 0318
- Clergé 0319
- Études bibliques 0321
- Histoire des religions 0320
- Philosophie de la religion 0322

Théologie 0469

SCIENCES SOCIALES
Anthropologie
- Archéologie 0324
- Culturelle 0326
- Physique 0327

Droit 0398
Économie
- Généralités 0501
- Commerce-Affaires 0505
- Économie agricole 0503
- Économie du travail 0510
- Finances 0508
- Histoire 0509
- Théorie 0511

Études américaines 0323
Études canadiennes 0385
Études féministes 0453
Folklore 0358
Géographie 0366
Gérontologie 0351
Gestion des affaires
- Généralités 0310
- Administration 0454
- Banques 0770
- Comptabilité 0272
- Marketing 0338

Histoire
- Histoire générale 0578
- Ancienne 0579
- Médiévale 0581
- Moderne 0582
- Histoire des noirs 0328
- Africaine 0331
- Canadienne 0334
- États-Unis 0337
- Européenne 0335
- Moyen-orientale 0333
- Latino-américaine 0336
- Asie, Australie et Océanie 0332

Histoire des sciences 0585
Loisirs 0814
Planification urbaine et régionale 0999
Science politique
- Généralités 0615
- Administration publique 0617
- Droit et relations internationales 0616

Sociologie
- Généralités 0626
- Aide et bien-àtre social 0630
- Criminologie et établissements pénitentiaires 0627
- Démographie 0938
- Études de l' individu et de la famille 0628
- Études des relations interethniques et des relations raciales 0631
- Structure et développement social 0700
- Théorie et méthodes. 0344
- Travail et relations industrielles 0629

Transports 0709
Travail social 0452

SCIENCES ET INGÉNIERIE

SCIENCES BIOLOGIQUES
Agriculture
- Généralités 0473
- Agronomie. 0285
- Alimentation et technologie alimentaire 0359
- Culture 0479
- Élevage et alimentation 0475
- Exploitation des péturages 0777
- Pathologie animale 0476
- Pathologie végétale 0480
- Physiologie végétale 0817
- Sylviculture et faune 0478
- Technologie du bois 0746

Biologie
- Généralités 0306
- Anatomie 0287
- Biologie (Statistiques) 0308
- Biologie moléculaire 0307
- Botanique 0309
- Cellule 0379
- Écologie 0329
- Entomologie 0353
- Génétique 0369
- Limnologie 0793
- Microbiologie 0410
- Neurologie 0317
- Océanographie 0416
- Physiologie 0433
- Radiation 0821
- Science vétérinaire 0778
- Zoologie 0472

Biophysique
- Généralités 0786
- Medicale 0760

SCIENCES DE LA TERRE
Biogéochimie 0425
Géochimie 0996
Géodésie 0370
Géographie physique 0368
Géologie 0372
Géophysique 0373
Hydrologie 0388
Minéralogie 0411
Océanographie physique 0415
Paléobotanique 0345
Paléoécologie 0426
Paléontologie 0418
Paléozoologie 0985
Palynologie 0427

SCIENCES DE LA SANTÉ ET DE L'ENVIRONNEMENT
Économie domestique 0386
Sciences de l'environnement 0768
Sciences de la santé
- Généralités 0566
- Administration des hipitaux 0769
- Alimentation et nutrition 0570
- Audiologie 0300
- Chimiothérapie 0992
- Dentisterie 0567
- Développement humain 0758
- Enseignement 0350
- Immunologie 0982
- Loisirs 0575
- Médecine du travail et thérapie 0354
- Médecine et chirurgie 0564
- Obstétrique et gynécologie 0380
- Ophtalmologie 0381
- Orthophonie 0460
- Pathologie 0571
- Pharmacie 0572
- Pharmacologie 0419
- Physiothérapie 0382
- Radiologie 0574
- Santé mentale 0347
- Santé publique 0573
- Soins infirmiers 0569
- Toxicologie 0383

SCIENCES PHYSIQUES
Sciences Pures
Chimie
- Genéralités 0485
- Biochimie 487
- Chimie agricole 0749
- Chimie analytique 0486
- Chimie minérale 0488
- Chimie nucléaire 0738
- Chimie organique 0490
- Chimie pharmaceutique 0491
- Physique 0494
- PolymÇres 0495
- Radiation 0754

Mathématiques 0405
Physique
- Généralités 0605
- Acoustique 0986
- Astronomie et astrophysique 0606
- Electronique et électricité 0607
- Fluides et plasma 0759
- Météorologie 0608
- Optique 0752
- Particules (Physique nucléaire) 0798
- Physique atomique 0748
- Physique de l'état solide 0611
- Physique moléculaire 0609
- Physique nucléaire 0610
- Radiation 0756

Statistiques 0463

Sciences Appliqués Et Technologie
Informatique 0984
Ingénierie
- Généralités 0537
- Agricole 0539
- Automobile 0540
- Biomédicale 0541
- Chaleur et ther modynamique 0348
- Conditionnement (Emballage) 0549
- Génie aérospatial 0538
- Génie chimique 0542
- Génie civil 0543
- Génie électronique et électrique 0544
- Génie industriel 0546
- Génie mécanique 0548
- Génie nucléaire 0552
- Ingénierie des systämes 0790
- Mécanique navale 0547
- Métallurgie 0743
- Science des matériaux 0794
- Technique du pétrole 0765
- Technique minière 0551
- Techniques sanitaires et municipales 0554
- Technologie hydraulique 0545

Mécanique appliquée 0346
Géotechnologie 0428
Matières plastiques (Technologie) 0795
Recherche opérationnelle 0796
Textiles et tissus (Technologie) 0794

PSYCHOLOGIE
Généralités 0621
Personnalité 0625
Psychobiologie 0349
Psychologie clinique 0622
Psychologie du comportement 0384
Psychologie du développement 0620
Psychologie expérimentale 0623
Psychologie industrielle 0624
Psychologie physiologique 0989
Psychologie sociale 0451
Psychométrie 0632

A REMOVAL SET BASED APPROACH TO RELIABILITY ASSESSMENT IN URBAN WATER SUPPLY DISTRIBUTION NETWORKS

BY

PAUL JACOBS

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Abstract

The exact calculation of water distribution system reliability has been shown to be NP-hard and, therefore, even for moderately sized systems, computationally impractical. Approximate methods to estimate water distribution system reliability are attractive if they can reduce the required computational effort. In this thesis a method is presented which can be used to estimate network reliability. The method is based on the definition of conditional probability, Monte Carlo simulation, and the partitioning of the failure space into disjoint sets to form a covering set.

The joint probability of system failure and the occurrence of a failure event in the covering set is calculated using the definition of conditional probability as the product of two terms. The first term is the probability of the failure of a specified number of components in one of the disjoint events in the covering set. This probability of occurrence is estimated using historical pipe failure data. The second term is the probability of system failure given that the same specified number of components have failed. This term is estimated using Monte Carlo simulation.

The total probability of system failure is estimated as the sum of the joint probabilities for all events in the covering set. Simultaneous estimates are required for both terms for each event in the covering set. Bonferroni techniques are used to perform the simultaneous estimation.

The technique is applied to four designs for a water distribution network. The total probability of system failure is estimated for all networks using both connectivity and hydraulic adequacy as failure criteria. Hydraulic adequacy is defined as the supply of adequate flow at sufficient pressure levels at all nodes in the network. The reliability of the network designs was successfully compared using both a confidence interval and a hypothesis testing approach.

Acknowledgements

This thesis would not have been possible without the help and support of a number of people. I would like to thank my advisor Professor Ian Goulter for his advice, and insightful comments. Professor Cass Booy also provided much help and encouragement in developing the statistics necessary for this thesis. I would also like to thank Jim Davidson for his introduction to and assistance with computer graphics. I also extend my thanks to Glen Chambers of the City of Winnipeg for his assistance with this thesis.

I am also indebted to my wife Maruta for her support and patience while I worked on this thesis.

Contents

1 INTRODUCTION **1**

2 Literature Review **5**

2.1 Network Design 5

2.2 Graph Theory 19

2.3 Estimating reliability: bounding the error 35

2.4 Solving a Hydraulic Network 40

3 Difficulties in the Exact Calculation of Network Reliability **44**

3.1 Comments 48

3.2 Use of Conditional Probability 49

4 A Simplified Model for Bounding Failure Probabilities **51**

4.1 Qualitative Basis of the Model 51

4.2 Theory 52

4.3 Error in the Estimate of the Probability of Network Failure 59

5 A Multiple Component Type Model **63**

5.1 Simultaneous Estimation 70

5.2 Confidence Limits for Total Probabilities 72

5.3 Factors Affecting Model Applicability 74

6 **ESTIMATING PROBABILITY OF EXACTLY m BREAKS** **77**

6.1 Sampling from the City of Winnipeg Data Base 78

6.2 Variation of Average Breakage Rates 83

6.3 Poisson Distribution 84

6.4 The Modified Poisson Model 91

6.4.1 Probability Distributions for Basic Events 93

6.4.2 Probability model for days with m=0 breaks 97

6.4.3 Probability model for days with m=1 break 98

6.4.4 Probability model for days with m=2 breaks 101

6.4.5 Probability model for days with m=3 breaks 104

6.5 Fitting the Modified Poisson Distribution 106

6.5.1 Estimation of LBI 107

6.5.2 Estimation of Parameter K0 108

6.5.3 Estimation of Parameter K1 111

6.5.4 Estimation of Parameter K2 115

6.5.5 Comparison of predicted to sampled probabilities 117

6.6 Confidence Limits for the Parameters LBI, K0, K1, and K2 119

6.7 Confidence Limits for Total Probabilities 119

6.8 Limitations of this Method 120

7 **Description of Model Networks** **121**

7.1 Assumed daily system operation plan 122

7.2 Example Networks 127

7.3 'Non-Optimal' system expansion specification 128

7.4 The Gessler network specification 131

7.5 Lee et al. solution specifications 133

7.6 Morgan and Goulter network specifications 135

7.7 Determining covering sets for pipe removal set probabilities 138

8 Estimation of Reliability Performance of Networks Using the Connectivity Criterion 156

8.1 Estimating the Probability of Node Disconnection 157

8.2 Rating the Solutions 168

9 Estimation of Reliability Performance of Networks using Hydraulic Criteria 174

9.1 Changes to the Monte Carlo Simulation Algorithm 175

9.2 The Non-Optimal Network 178

9.3 The Gessler Network 182

9.4 The Lee et al. Network 182

9.5 The Morgan and Goulter Network 185

9.6 Discussion of Results 185

9.7 Computational Requirements 193

10 Discussion of Results 196

10.1 Evaluation of Solution Performance 196

10.2 Failure Criteria 197

10.3 Sources of Error 198

10.4 Effects of System Size and Complexity 200

10.5 Mean Time to Repair 201

11 Summary and Conclusions 203

A Data Transformation and Verification 213

A.1 Description of the Pipe Break and Pipe Inventory Data 213

A.2 Techniques to Create the Pipe Inventory 218

B Additional Figures to Estimate Component Failure Probabilities 228

List of Figures

2.1 Example of a Series Reduction 22

2.2 Example of a Parallel Reduction 22

3.1 Example of a Tree Network 46

3.2 Example of a Series Parallel Reducible network 48

6.1 Schematic of Computer System to Extract Pipe Breakage Samples . . 80

6.2 Poisson estimated probability for m=0 breaks: 150mm Cast iron, 0-19 years 86

6.3 Poisson estimated probability for m=1 break: 150mm Cast iron, 0-19 years 87

6.4 Poisson estimated probability for m=2 breaks: 150mm Cast iron, 0-19 years 88

6.5 Poisson estimated probability for m=3 breaks: 150mm Cast iron, 0-19 years 89

7.1 Original Network Layout from Walski, 1987 122

7.2 Non-Optimal Network Layout 131

7.3 Gessler Network Layout 132

7.4 Lee et al. Network Layout 135

7.5 Morgan and Goulter Network Layout 138

A.1 Schematic of Computerized Map Correction System 220
A.2 Segment of Computerized Map Before Validation 222
A.3 The Same Segment of the Computerized Map after Correction 227

List of Tables

6.1 Average number of breaks per kilometer per year for the City of Winnipeg 79
6.2 Predicted number of days with m independent breaks for 35 km of 150 mm cast iron pipe older than 30 years 96
6.3 Estimation of parameter LBI 109
6.4 Estimation of parameter KDD 112
6.5 Summary of value of KDD, LBI, and K0 113
6.6 Estimation of Parameter K1 116
6.7 Estimation of Parameter K2 118

7.1 Fixed data for links for all networks 123
7.2 Node characteristics table 124
7.3 Link data for original network 125
7.4 Daily demand pattern (average = 1.0) 126
7.5 Data for changed links in the non-optimal network 129
7.6 Specification of tanks and pumps - non-optimal network 130
7.7 Data for changed links in the Gessler network 133
7.8 Specification of tanks and pumps - Gessler network 134
7.9 Data for changed links in the Lee et al. network 136
7.10 Specification of tanks and pumps - Lee et al. network 137
7.11 Data for changed links in the Morgan and Goulter network 139

7.12 Specification of tanks and pumps - Morgan and Goulter network . . . 140
7.13 Pipe categories and abbreviations 141
7.14 Pipe length by pipe category - non-optimal network 142
7.15 Pipe length by pipe category - Gessler network 142
7.16 Pipe length by pipe category - Lee et al. network 143
7.17 Pipe length by pipe category - Morgan and Goulter network 143
7.18 Bonferroni confidence limits for component failures - non-optimal network 144
7.19 Bonferroni confidence limits for component failures - Gessler network 145
7.20 Bonferroni confidence limits for component failures - Lee et al. network 146
7.21 Bonferroni confidence limits for component failures - Morgan and Goulter network 147
7.22 Covering set - days with exactly 0 or 1 break: non-optimal network . 149
7.23 Covering set - days with exactly 2 breaks: non-optimal network . . . 149
7.24 Covering set - days with exactly 3 breaks: non-optimal network . . . 150
7.25 Covering set - days with exactly 0 or 1 break: Gessler network 150
7.26 Covering set - days with exactly 2 breaks: Gessler network 151
7.27 Covering set - days with exactly 3 breaks: Gessler network 151
7.28 Covering set - days with exactly 0 or 1 break: Lee et al. network . . . 152
7.29 Covering set - days with exactly 2 breaks: Lee et al. network 152
7.30 Covering set - days with exactly 3 breaks Lee et al. network 153
7.31 Covering set - days with exactly 0 or 1 break: Morgan and Goulter network 153
7.32 Covering set - days with exactly 2 breaks: Morgan and Goulter network 154
7.33 Covering set - days with exactly 3 breaks Morgan and Goulter network 154
7.34 Confidence limits for cumulative probabilities - all networks 155

8.1 Conditional failure probabilities - days with exactly 0 or 1 break: non-optimal network 158
8.2 Conditional failure probabilities - days with exactly 2 breaks: non-optimal network 159
8.3 Conditional failure probabilities - days with exactly 3 breaks: non-optimal network 159
8.4 System failure probabilities - days with exactly 0 or 1 break: non-optimal network 160
8.5 System failure probabilities - days with exactly 2 breaks: non-optimal network 160
8.6 System failure probabilities - days with exactly 3 breaks: non-optimal network 161
8.7 Conditional failure probabilities - days with exactly 0 or 1 break: Gessler network 161
8.8 Conditional failure probabilities - days with exactly 2 breaks: Gessler network 162
8.9 Conditional failure probabilities - days with exactly 3 breaks: Gessler network 162
8.10 System failure probabilities - days with exactly 0 or 1 break: Gessler network 163
8.11 System failure probabilities - days with exactly 2 breaks: Gessler network 163
8.12 System failure probabilities - days with exactly 3 breaks: Gessler network 164
8.13 Conditional failure probabilities - days with exactly 0 or 1 break: Lee et al. network 164
8.14 Conditional failure probabilities - days with exactly 2 breaks: Lee et al. network 165

8.15 Conditional failure probabilities - days with exactly 3 breaks: Lee et al. network 165
8.16 System failure probabilities - days with exactly 0 or 1 break: Lee et al. network 166
8.17 System failure probabilities - days with exactly 2 breaks: Lee et al. network 166
8.18 System failure probabilities - days with exactly 3 breaks: Lee et al. network 167
8.19 Conditional failure probabilities - days with exactly 0 or 1 break: Morgan and Goulter network 169
8.20 Conditional failure probabilities - days with exactly 2 breaks: Morgan and Goulter network 169
8.21 Conditional failure probabilities - days with exactly 3 breaks: Morgan and Goulter network 170
8.22 System failure probabilities - days with exactly 0 or 1 break: Morgan and Goulter network 170
8.23 System failure probabilities - days with exactly 2 breaks: Morgan and Goulter network 171
8.24 System failure probabilities - days with exactly 3 breaks: Morgan and Goulter network 171
8.25 Confidence intervals for cumulative probability of system failure . . . 172
8.26 Comparison of Morgan and Goulter network to other networks 172

9.1 Conditional failure probabilities - days with exactly 0 or 1 break: non-optimal network 179
9.2 Conditional failure probabilities - days with exactly 2 breaks: non-optimal network 179

9.3 Conditional failure probabilities - days with exactly 3 breaks: non-optimal network 180
9.4 System failure probabilities - days with exactly 0 or 1 break: non-optimal network 180
9.5 System failure probabilities - days with exactly 2 breaks: non-optimal network 181
9.6 System failure probabilities - days with exactly 3 breaks: non-optimal network 181
9.7 Conditional failure probabilities - days with exactly 0 or 1 break: Gessler network 182
9.8 Conditional failure probabilities - days with exactly 2 breaks: Gessler network 183
9.9 Conditional failure probabilities - days with exactly 3 breaks: Gessler network 183
9.10 System failure probabilities - days with exactly 0 or 1 break: Gessler network 184
9.11 System failure probabilities - days with exactly 2 breaks: Gessler network 184
9.12 System failure probabilities - days with exactly 3 breaks: Gessler network 185
9.13 Conditional failure probabilities - days with exactly 0 or 1 break: Lee et al. network 186
9.14 Conditional failure probabilities - days with exactly 2 breaks: Lee et al. network 186
9.15 Conditional failure probabilities - days with exactly 3 breaks: Lee et al. network 187
9.16 System failure probabilities - days with exactly 0 or 1 break: Lee et al. network 187

9.17 System failure probabilities - days with exactly 2 breaks: Lee et al. network 188

9.18 System failure probabilities - days with exactly 3 breaks: Lee et al. network 188

9.19 Conditional failure probabilities - days with exactly 0 or 1 break: Morgan and Goulter network 189

9.20 Conditional failure probabilities - days with exactly 2 breaks: Morgan and Goulter network 189

9.21 Conditional failure probabilities - days with exactly 3 breaks: Morgan and Goulter network 190

9.22 System failure probabilities - days with exactly 0 or 1 break: Morgan and Goulter network 190

9.23 System failure probabilities - days with exactly 2 breaks: Morgan and Goulter network 191

9.24 System failure probabilities - days with exactly 3 breaks: Morgan and Goulter network 191

9.25 Confidence intervals for cumulative probability of system failure . . . 192

9.26 Comparison of Morgan and Goulter network to other networks 193

9.27 Elapsed running times and associated errors for Monte Carlo Simulation Runs 194

A.1 LISP Macros and Their Functions 224

Chapter 1

INTRODUCTION

A number of approaches to the design of water distribution networks using mathematical programming models have been attempted in the literature. This design problem has been well solved for the case in which only cost criteria is considered. Networks designed using least cost criteria have not, however, generally proven acceptable from a reliability point of view because such networks are branched. Branched networks which represent the cheapest systems are considered unacceptable for urban water supply purposes because they do not provide the necessary alternate supply paths, should a link fail. In addition, such networks contain dead end links, in which water can stagnate. Such stagnation promotes the growth of algae and bacteria and, hence, can lead to a health hazard.

The design of water distribution system models,which include reliability criteria in a satisfactory manner has proven to be a difficult task. The difficulty of this design task has been increased by the lack of a clear definition of system reliability. Both heuristic and probabilistic definitions of system reliability have been proposed.

The major advantage of heuristic measures is the relatively modest computational burden required for their use. Several authors have incorporated heuristic definitions of reliability into design models (Elms (1983), Morgan and Goulter (1985), Coals and Goulter (1985), Awumah et al. (1991)). Modifying models currently in use for design of non-branched networks by the inclusion of heuristic reliability measures has also been attractive (See Walski et al., 1987). The results have been interesting, but have not been widely employed. Furthermore, the different design methods produce different results.

All heuristic based methods currently found in the literature use surrogate measures for network reliability. They provide only a relative and imperfect indication of reliability. Furthermore, most design methods do not define the events that constitute a network failure explicitly. Hence, it is generally impossible to quantitively compare designs that arise from the different approaches.

These problems associated with heuristic reliability measures make the idea of a model based on probabilistic reliability measures attractive. The calculation of network reliability has, however, proven to be a difficult and computationally intensive problem. Ball (1980) showed that the problem is NP hard. By adding 3 links to a 14 link network and increasing computation time by a factor of 10 Su et al. (1987) demonstrated that these reliability calculations are computationally impractical. Unless difficulties of calculating reliability can be overcome, the development of a design model based on probabilistic reliability measures is not possible.

One approach to the reduction of the computational burden associated with the use of probabilistic methods is to *estimate* the probabilities used to determine system reliability by sampling. Ball and Provan (1983) showed that, under the severely restrictive and non-conservative assumption that all links in a network have the same probability of failure, it is possible to approximate reliability to any desired level of accuracy. This approximation is based the derivation of polynomial expressions to calculate the probability of 0, 1, 2, 3, etc. simultaneous failures. Ball and Provan's approach also assumed that network failure occurs only if at least one node is isolated, neglecting the possibility of failure in the form of inadequate water pressure. This approach has not been applied to the design of water distribution systems for this reason and because the underlying assumption of equal probability of failure is false for most practical networks. Network failure can also arise from either inadequate water pressure at the demanded flow rates or the delivered flow rates can be inadequate to meet the required demand.

Despite the limitations of Ball and Provan's approach, the concept of estimating network failure probabilities using failure probability polynomials is attractive because it would provide an approximate, explicit rather than surrogate, probabilistic measure of reliability. Ball and Provan assume that all links fail independently and have the same probability of failure. This assumption is not supported by the information about water distribution networks. It should also be noted, that in general, much more information exists about pipe failure in water distribution networks than is used in Ball and Provan's approach. In particular, the City of Winnipeg has kept extensive, computer readable records of pipe failures. These records provide information about the number of simultaneous failures that are likely and about the failure pattern that has occurred in the past. In this thesis, a method is proposed to use additional data similar to those in the City of Winnipeg break data base to estimate pipe network reliability without an excessive computational burden.

The network failure probabilities considered in the methodology being proposed in this thesis can be decomposed into two components. The first component is the probability of exactly m simultaneous breaks occurring (m = 0, 1, 2, 3, etc.) It is independent of network topology. These probabilities can be estimated from the break records. The second component is the probability of network failure, given that exactly m simultaneous breaks have occurred. It is dependent on the network topology. This conditional probability can be estimated using Monte Carlo Simulation methods. Two different criteria will be used to determine network failure. These criteria are:

- node disconnection
- inadequate pressure or flow levels at the demand nodes in the network

In the next chapter a detailed review is given of approaches taken to the design of optimally reliable water supply systems, in the literature.

Chapter 2

Literature Review

This review is divided into four sections representing the four different types of research relevant to the study reported in this thesis. The first area of research is primarily concerned with the development of models for optimal (least cost) design of water distribution systems. The second area of research reviewed is primarily concerned with the application of graph theory concepts to the evaluation of network reliability. The third area of research is concerned with the use of Monte Carlo simulation to estimate network reliability. The fourth area of research is specifically focused on the solution (pressure and flow determination) of a hydraulic network which is a key component of the simulation program used in this study.

2.1 Network Design

Design of optimal water distribution networks, Karmeli et al. (1968).

In this article one of the first linear programming models developed to design optimal water distribution systems is presented. The cost function is linearized by using lengths of pipe of standard diameters as the decision variables. The non-linearities associated with head loss are linearized by noting that, for a specific link of specific diameter, the head loss, per unit length, is a constant. Using these assumptions, a linear programming model to optimize a branched water distribution network with a single source of supply is developed. No attempt is made to include reliability considerations in the model.

Optimal design and operation of water distribution systems, Shamir (1974)

In this article a non-linear model to address the optimal design and operation of water distribution systems is developed. An interesting aspect of the formulation is the recognition that it is necessary to constrain either pressure head gradients between nodes or flows between nodes, but not both. The model handles looped networks as well as branched networks. This model is a significant advance over the model of Karmeli et al. (1968) because current design practice requires that urban water supply networks contain loops. However, no specific recognition is made of reliability considerations and the model suffers from problems inherent in a non-linear programming approach. These problems include the inability to guarantee that a global optimum is found, the requirement for manual intervention to prevent non-convergence of the model and the possibility of a large computational burden to produce a solution.

Design of optimal water distribution systems, Alperovits and Shamir (1977)

In this article several improvements are made relative to the model previously proposed by the second author. Rather than using a non-linear programming model to solve the water distribution system design problem, an iterative approach using a linear programming model at each step is employed. In the model flows are assumed to be fixed for each step (solution of the linear program). These flows are adjusted by adjusting the flows in the pipes on the basis of the gradient of the network cost with respect to the flows. This model was a landmark in the methodology and has become a standard for the solution of the water distribution network design problem in which loops exist and for which network reliability considerations are recognised only by ensuring that loops exist in the network.

The authors recognized this limitation explicitly and called for a more intrinsic definition of system reliability than the requirement that the network contain redundancy in the form of loops. Such a definition of an intrinsic and practically applicable measure of system reliability has not been accomplished to date.

Optimization of looped water distribution systems, Quindry et al. (1981)

In this article another linear programming gradient model is developed to optimize a water distribution system which contains loops. The approach is similar to that used by Alperovits and Shamir (1977), the main difference between these two methods being that the Quindry et al. (1981) model assumes fixed heads and, based on the assumed heads, solves a linear programming model, for the optimum flows in each link. After each iteration, the heads are changed in the direction of the gradient of the cost with respect to the assumed heads. In contrast, Alperovits and Shamir (1977) assumed flows and modified the heads in the direction of the gradient, with respect to cost.

The formulation, presented in this article is simpler and has lower computational requirements than the model used by Alperovits and Shamir (1977). The authors also recognize that the requirement for redundancy provided by loops in the network is vague, but do not directly address the issue of the design of optimally reliable water distribution networks.

It is interesting to note that in the solution to the New York City supply tunnels problem, the solution found by the Quindry et al. (1981) model, was an implicitly branched network. All existing pipes in the system were assumed to remain in the system, with all expansion being concentrated on a spanning tree of the network.

Obtaining layout of water distribution systems, Rowell and Barnes (1982)

In this article the authors develop a two part model to design the layout and size the elements of a water distribution network. The first part of the model selects a tree shaped network to provide basic flow to demand nodes. The second part of the model adds links to form loops in the tree structure. These loops are identified as those necessary to provide alternate paths from the source node to the demand nodes in case of primary link failure and system maintenance.

The first part of the model is a non-linear programming model which minimizes the cost of providing pipes for a branched network subject to linear flow constraints. The conditions under which the cost function is concave are noted and it is shown that, for normal operating conditions for water supply systems, the cost function meets these conditions. Since the model involves minimizing a concave cost function over a linear set of constraint equations, this model guarantees a globally optimal solution for the branched network under consideration.

The second part of the model, which generates redundant links, is a general {0-1} integer programming problem, which minimizes the cost of providing redundant links subject to the requirement that the flow capacity of the redundant links is greater than a specified minimum value. The authors note that, in general, this approach does not provide an optimal design for the total network.

This two part model is of interest because it is the first model in the literature to partition the design of water supply networks into two parts and to use the output of the first part of the model as input to the second part. It should be noted that the assumption implicit in this approach is that the provision of at least two independent paths to each node ensures sufficient reliability for the network.

In addition, a constant hydraulic gradient is assumed for all links in the network. Goulter and Morgan (1984) identified this as a problem in maintaining hydraulic consistency in this model.

Discussion of 'Obtaining layout of water distribution systems' by Rowell and Barnes, Goulter and Morgan (1984)

In this discussion paper the authors point out that the assumption of Rowell and Barnes that the head loss gradient is constant is only reasonable in a tree shaped network. In a looped network hydraulic consistency requires that the head loss between two points have the same value, regardless of path. If paths are of different lengths, the head loss gradient must vary across the network.

A feedback mechanism, connecting the two parts of the model, is suggested to ensure that the model better represents the actual conditions that would be present in a functioning system. It should be noted that such a feedback mechanism implies that the two parts of the model would be iterative in nature. This would impose a larger computational burden than that reported by Rowell and Barnes (1982).

An integrated approach to the layout and design of water distribution networks, Goulter and Morgan (1985)

In this article the design of water distribution networks is again separated into two sub-problems: network layout and pipe sizing. A linear programming model (a pressure head model) based on assumed flow patterns is proposed to solve the pipe sizing problem. Another linear programming model is proposed to solve the layout problem (the layout model) based on assumed pressure heads. Initially, pressure heads are assumed for the layout model. The output of the layout model includes generated flows, based on the assumed pressure heads. These flows are used as input to the pressure head model. The pressure head model, in turn, generates revised pressure heads. This iterative procedure is repeated until a satisfactory design is achieved.

This design method equates network reliability with the existence of redundant paths (loops) to each node. The model attempts to force redundant paths into networks designed by this model by means of a condition which guarantees that two links are incident on each node. The authors note that this condition is insufficient to guarantee true redundancy. Visual inspection of the network is required to detect the existence of 'non-redundancies' in the network that this reliability condition fails to preclude. Such 'non-redundancies' includes no true alternative pathway from a source to a demand point. In addition, manual intervention, which can be automated, is necessary between successive runs, of each model, to ensure that reliability conditions are being met.

Quantitative approaches to reliability assessment in pipe networks, Goulter and Coals (1986)

This paper contains a general discussion of reliability criteria for water distribution networks. Network reliability is designed as the probability of the network being able to supply sufficient flow at the required pressure to all nodes in the network. The paper recognizes that both the network layout and the component reliabilities contribute to network reliability. Specifically, it is recognized that a single pipe which has a very low probability of failure might provide a higher level of reliability than several less reliable pipes connected in parallel. The separation of component reliability from layout reliability is a key component of the model presented in this thesis.

In addition, two methods of designing an optimal water distribution system under reliability constraints are proposed. The first method used a node isolation approach. In this approach the probability that all links connecting a node to the network are broken is used as a measure of reliability. This measure of reliability is non-conservative because it is possible to isolate a node from the network by breaking links that are not directly connected to the node. In this sense this model uses a proxy for reliability rather than measuring reliablity directly.

The second model uses a goal programming approach based on minimizing the differences in reliabilities of pipes connected to each node. It should be noted that this approach does not explicitly take into account component reliability and as in the previous method is non-conservative because only links directly connected to a node are included in the goal programming formulation. In effect, this model uses a proxy measure for layout reliability.

This article is interesting because it proposes a clear definition of network reliability. There are problems, however, in the application of both reliability proxies, such as those in this study, because none of them can guarantee that a network design will work with a given probability level.

Optimal urban water distribution design, Morgan and Goulter (1985)

In this article a heuristic, iterative model for the design of urban water distribution networks is presented. A linear programming model is used to determine the optimal links in a network in which to change pipe diameters. The Hardy-Cross method is used then to check for hydraulic consistency. A linear weighting algorithm is used to decide which pipes are to be eliminated from the network after each iteration.

This article is interesting because it is the first article in the literature to distinguish between network reliability and the design of looped networks. The model developed does not explicitly include constraints to enforce any reliability proxy. Instead, redundancy is forced by the consideration of a large number of test cases (load patterns). These test cases included normal usage demands at all nodes, except one, and the worst possible pipe break, combined with the demand for the simultaneous provision of fire flow volumes at the remaining node. One test case was specified for each demand node. The use of a large number of test load cases to force redundancy into the network, is a useful extension to the concept that network reliability is essentially the ability to provide for all required demands at all nodes of the network when a component in the network, for example a pipe, is broken.

The method proposed in this article is heuristic in nature and, hence, does not guarantee a globally optimal solution. Again because this method determines network reliability qualitatively, rather than numerically, the results produced by this model are not directly comparable to other results produced by other models.

Reliability-based optimization model for water distribution systems, Su et al. (1987)

In this article a three part model is proposed to analyze and design optimally reliable water supply systems. The model components are an optimization model, a system simulation model, and a system reliability model.

System reliability is calculated using the minimum cut-set method in which cut-sets are limited to the simultaneous failure of two links. Every possible two link cut-set is tested by removing a pipe from the system and testing the resulting system using KYPIPE (Wood, 1980), a water distribution system simulation model to test for inadequate pressure heads. If inadequate pressure heads are found at any demand node, then a system failure has been identified. When all single pipes have been checked in this way, combinations of two pipes are checked.

The optimization model uses the generalized reduced gradient method GRG2 to solve and optimize the pipe network. Problems were encountered in defining the gradient of maximum improvement in system reliability. No analytical method was able to be formulated to compute this gradient. A numerical scheme was developed that perturbs each decision variable individually and computes the objective function and constraints for the perturbed values and then calculates the required gradients.

This model was run on two example distribution systems. The first system had 14 links and 1 loop. The execution time required to solve this system was 1157 seconds on a dual CDC cyber 170/750 mainframe. The second system had 3 additional pipes and 4 loops. Execution time jumped to 200.5 minutes (12,030 seconds). The above discussion makes it clear that this model is computationally intensive to an extreme degree and that it becomes very much more computationally intensive as system size, and, in particular, the number of loops (redundant paths), increases to realistic levels.

The computational intensity of this model is partially due to the non-linear programming model used to optimize the system and to the large number of removal sets that had to be tested for each iteration despite the arbitrary limit of two for the cut-set size. The lack of an analytical method to determine the gradient of reliability with respect to network flows is another factor which added greatly to the computational burden associated with this method.

This article is interesting because it confirms graph theoretical results that the reliability problem is computationally intensive. These graph theoretic issues are discussed in more detail in Section 2.2. It is also interesting,however, to note that very little difference in reliability was found when reliability was tested using single link cut sets, as opposed to two link cut sets. This result corresponds well with the result obtained by Jacobs and Goulter (1991), that for small networks, a cut-set size of 1 is sufficient to achieve a good estimate of reliability if links are assumed to fail independently. The idea of using preset pressure heads and flows as design criteria also provides a clear and measurable criterion for system failure. The use of a hydraulic network solver program to test for system failure provided a feasible method to determine the applicability of this criterion for system failure.

A modified linear programming gradient method for optimal design of looped water distribution networks, Fujiwara et al. (1987)

In this article an iterative linear programming model to optimize the design of a looped water supply network based only on cost considerations is presented. The basic linear programming model is similar to that used in Morgan and Goulter (1985) and was originally presented in Alperovits and Shamir (1977). A formal expression is identified for the gradient of the cost function with respect to changes in flow along each flow path. The computational efficiency of the original Alperovits and Shamir (1977) model is improved by using backtracking and quasi-Newton search and convergence methods, rather than the steepest gradient method.

The authors found a lower cost solution to the problem originally solved by Alperovits and Shamir (1977). They also found that the solution is quite sensitive to search step size and starting solution. It was speculated that the solution cost function has many local optima within close proximity to each other.

Despite the inclusion of loops in the water supply systems modeled, no attempt is made to specifically include network reliability in the model. Given the difficulties encountered by Su et al. (1987) in determining a gradient function for cost of a system with respect to system reliability, it seems doubtful that the method could, by itself, be generalized to handle reliability constraints. The method does, however, have promise as part of an iterative approach to solving the reliability problem.

Battle of the network models: epilogue, Walski et al. (1987)

In this article a representative water distribution network design problem was solved by six groups each using their own model. All solutions were found to be close in cost. The most interesting aspect of this article from the point of view of the problem examined in this thesis, is that it was impossible to declare a 'winner' in the contest. It was suggested that the costs of the solutions generated by all models are similar and that differences were mainly attributable to the location of storage tanks. Interestingly, the most expensive solution was, in fact, that having the greatest storage which has the implication of the solution being the most reliable system.

No attempt was made to evaluate the relative reliability of the solutions. However, the desirability of having a more precise definition of reliability requirements that would make the results of optimization models more meaningful was recognized.

Network reliability for water distribution system management Quimpo and Shamsi (1990)

In this article an interesting graphic heuristic to aid in decision making concerning network reliability is proposed. The graphic is based on the S-T (source to terminal) connectivity measure of reliability. The S-T measure determines the probability that a given node is connected to the source node. In this paper the S-T reliability is calculated for all demand-source node combinations in the network. The reliabilities are plotted for all nodes and lines of equal probability are drawn based on the nodal reliabilities. The resulting plot can then be used to identify areas of the network in which reliability targets are not met.

The paper uses inclusion-exclusion formulas to estimate network reliability. It should be noted that in the worst case the use of these formulas requires the evaluation of an exponential number of terms with respect to the number of components in the network. Consequently, truncation of the series, when a required level of accuracy is reached is suggested. No method is given to determine the number of terms that are required to reach the required level of accuracy. The paper does not adequately address this issue. In this thesis, for conditions in the City of Winnipeg, it is shown that two or at most three simultaneous breaks need be considered to adequately estimate reliability.

In addition, it should be noted that the method does not explicitly calculate or estimate network reliability. If two alternate designs for a network are considered it is possible that some parts of the network might have higher levels of reliability under the first design, while other parts of the network have higher levels of reliability under the second design. This limitation can lead to difficulties in the comparison of the reliability of alternate system designs.

An important aspect of this article from the point of view of this thesis is that it recognizes that network reliability is influenced both by the reliability of individual network components and the layout of those components.

Computerized evaluation of water-supply reliability Shamsi (1990)

In this paper a more detailed explanation of the technique proposed in Quimpo and Shamsi (1990) is given. Detailed explanations of the computer programming and graph theory techniques necessary to employ this method are provided. A three-dimensional plot of reliability versus geographical position is again drawn.

The component reliabilities are calculated using the inclusion-exclusion formulae for either pathsets or cutsets. In general, both of these formulas have an exponential number of terms, with respect to the number of pipes in the network. It is noted that in systems with low reliability components the minimal pathset analysis may not satisfy the accuracy criterion and full enumeration of all terms may be required.

In summary, the graphical technique presented in this article is interesting, but it does not address any of the difficult computational problems pertaining to reliability calculations that have been previously raised in the literature.

Reliability-Based Distribution System Maintenance Quimpo and Shamsi (1991)

In this paper the technique described in Quimpo and Shamsi (1990) and Shamsi (1990) is extended and refined. The technique is applied to several networks including two real networks, the Braddock Water Authority in Pennsylvania and the City of Norwich in New York. The City of Norwich network is relatively large containing 126 nodes, 192 arcs, and 25 loops. The technique was applied successfully to create a reliability surface for this network. It should, however, be noted that for this network the ratio of loops to arcs is relatively low as compared to the Walski et al. (1987) network. The ratios are 0.13 for the City of Norwich network versus 0.49

for the Walski et al. (1987) network. It is possible that the low ratio of loops to arcs found in the City of Norwich network could enhance effectiveness of the series-parallel reducitons used to reduce the complexity of calculating network reliability. More work is required to determine the effect of increasing the loop to arc ratio on the computational effort required to implement the procedure presented in this paper. It is impossible to compare the computational effort required to implement the method presented in the paper with other methods presented in the literature because the computational effort required is not presented.

The paper recognized the importance of pipe material, pipe age, and pipe diameter in predicting component reliability but assumed the component reliabilities. The existence of failure history data was noted, but such failure history data were not used to calculate component reliabilities. The procedure proposed in this thesis makes explicit use of these historical data base parameters to predict component failure probabilities.

Problems with the use of connectivity as a criteria for network failure are also recognized. The possibility that component failures might result in inadequate pressure or flow levels without disconnecting the network is recognized. This weakness in the model presented in the paper is identified as an area for future research. It should be noted that this thesis introduces techniques to use hydraulic adequacy, namely, the provision of adequate flow and pressure levels in the network, as the failure criterion.

This paper is especially interesting because of the clear decomposition of network relibility into component reliability and network topology. This concept forms the basis of the formal decomposition used in this thesis to estimate network reliability.

Predictive model and reliability analysis for water distribution systems
Wu and Quimpo (1992)

In this paper the methods presented in Quimpo and Shamsi (1990), Shamsi (1990), and Quimpo and Shamsi (1991) are extended. The use of a proportional hazards model is suggested to predict component reliabilities. Details concerning the application of the proportional hazards model are not provided in the paper. Furthermore in the case study presented the proportional hazards model is not applied and values for component reliabilities are assumed.

Three types of reductions are employed to reduce the computational effort required to calculate reliability. These reductions are series reductions, parallel reductions, and type 1 polygon-to-chain reductions. An important innovation is the use of these reductions to calculate both expected probability of failure and capacity for reduced links. The paper claims that the use of these reductions removes the reliability problem from the NP-hard class. This claim is false for general networks because all networks are not completely reducible using the reductions given in this paper. When all possible reductions have been made, an unreduced portion of the network may remain. If this situation occurs, the paper suggests using the factoring theorem to further reduce the network. In the worst case employment of the factoring theorem requires 2^{n-1} operations (n = the number of links in the network).

Despite these theoretical problems the use of the type 1 polygon-to-chain reduction is valuable because it allows reduction of some networks that can not be reduced using series or parallel reductions. In addition, the calculation of expected reduced link capacities makes it possible to use adequacy of flow as a failure criterion. The paper shows that reliability calculated using adequacy of flow as the failure criterion is always less than or equal to reliability calculated using connectivity as the failure criterion for a small sample network. This result suggests that the use of connectivity as a failure condition is non-conservative. The paper does not attempt to use adequacy of pressure as a failure criterion and all reliability calculations are related to the problem of source to terminal connectivity.

2.2 Graph Theory

Matroids and a reliability analysis problem, Ball and Nemhauser (1979).

In this article reachability is defined as the probability that all demand nodes are connected to a single source node. The reachability problem is formulated in terms of stochastic coherent binary systems. The basic building blocks of a stochastic coherent binary system, in the reliability context, are tuples of the form $= (S_1, S_2, ..., S_m)$. Each S_i corresponds to a component of the system under consideration and can take on one of two values, depending on whether the component is working or not. The system failure set is then defined as the set of tuples that cause the system to fail. Uniform probabilities of failure are assumed for all system components. A binomial model is defined to evaluate these probabilities.

The authors show that the calculations required by this model are exponential, with respect to the number of components in the system. A method is then presented to decompose the system to process larger groups of tuples at a time and thus reduce the computational burden. While the method does reduce computations somewhat, the number of computations required still increases exponentially with system size.

A recursive algorithm for finding reliability measures related to the connection of nodes in a graph, Buzacott (1980)

In this article a recursive algorithm is presented to calculate the reliability of a network, with known probabilities of failure for each link. The algorithm works by recursively decomposing the network into subsets. This algorithm appears to be the fastest of the methods for solving the reachability problem that have appeared in the literature. The computational requirements are of the order $3^n - 1$ where n is the number of links in the network and are demonstrated in the article only for undirected networks .

The article also contains a summary of the approaches taken to solving the reachability problem and the associated computational requirements.

The complexity of network reliability computations, Ball (1980)

In this article a number of reliability analysis problems are examined to determine the computational burden associated with their solution. Of particular interest for this study is the reachability problem for networks with nodes that can fail. This case considers the connection of a source node to all other nodes of the network. Since the basic (necessary but not sufficient) criterion for urban water supply network reliability is that all demand nodes are still connected to at least one source node when one or more links have failed, reachability appears to be the most appropriate model to use in evaluation of water supply network reliability.

It was found that the reachability problem with unreliable nodes is NP-hard. In other words, the number of computations required to solve the reliability problem increases exponentially with the number of links and nodes in the network. In essence, the paper shows that, while these problems are soluble, the computational burden of solving them is excessive for all but trivially small problems.

Series-parallel reduction for difficult measures of network reliability, Rosenthal (1981)

The previous articles from graph theory, in particular Ball (1980), show that, in the general case, the calculation of network reliability is NP-hard. This article presents the series-parallel method for the reduction of the calculations required to determine network reliability.

The reliability problem is presented as a network problem in which the links of a network G can be working or not working. The probability of the network working is then presented as the product of the probabilities that each link necessary to perform a specific task is working. It should be noted that this method assumes that link failure probabilities are independent of each other.

The method works by decomposing the network into subsets. These subsets are treated as complex edges in a summarized network. A complex edge is an edge that has more than two possible states. More than two possible states can occur in a complex edge because each link in the subset represented by the complex edge can be either working or failed. Each combination of failed and working links then represents a unique value for the complex edge. For example, consider a subset of a network that is made up of three links in parallel. The possible states in this case are all links working, all combinations of two links working, the three cases in which only one link works, and the case in which all links are not working.

These subsets are chosen so that the reliability of each subset can be calculated, independently of the rest of the network. In general terms the subsets are chosen so that:

1. subset reliability is calculable.

2. if two subsets, E_1 and E_2, of the network are chosen then sufficient information must be stored about these subsets to calculate the reliability of $E_1 \cup E_2$.

3. the reliability of a new network that treats the subsets as complex edges can be calculated without reference to the original network.

The effect of partitioning a network according to the above criteria is to simplify the calculation of network reliability by replacing subsets of the original network containing multiple links and nodes, with corresponding complex edges, each containing a single link and two nodes. It should be noted that it is not always possible to partition a network according to the above conditions.

Two specific methods by which networks can be partitioned in accordance with the above conditions are presented. These are series reductions and parallel reductions. A series reduction replaces several links connected, in series, with a single complex edge (See Fig. 2.1). A parallel reduction replaces two disjoint paths with common source and terminal nodes with a complex edge (See Fig. 2.2).

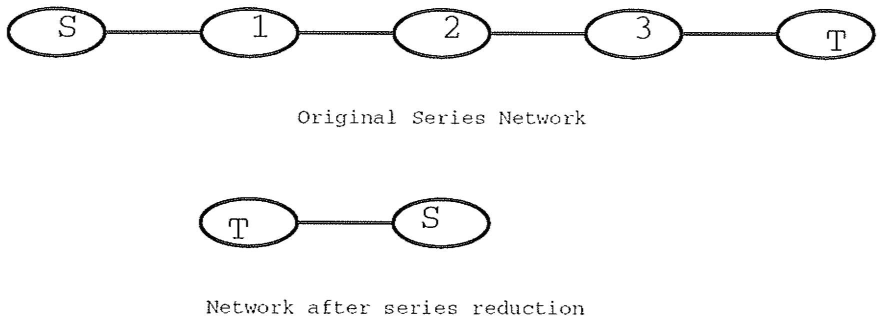


Figure 2.1: Example of a Series Reduction

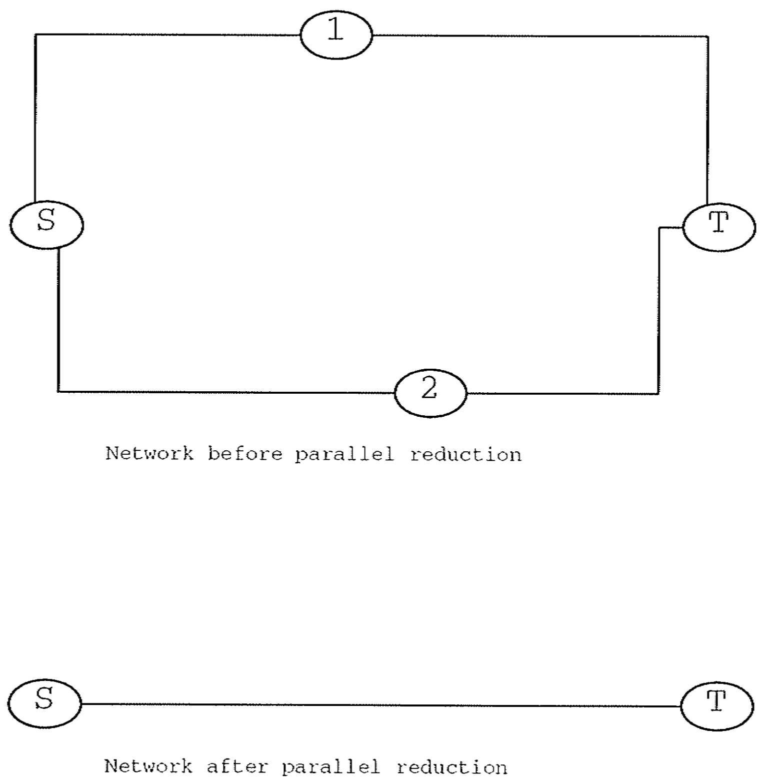


Figure 2.2: Example of a Parallel Reduction

By successive applications of this method it is possible to simplify a network to make the reliability calculation more tractable. It is noted, however, that it is possible to exhaust all possibilities for simplifications before the network has been reduced to a state for which reliability is easily calculable. A number of papers in the literature, e.g., Shamsi (1990), Quimpo and Shamsi (1990), Quimpo and Shamsi (1991), and Awumah and Goulter (1991), have applied series and parallel reductions to calculate network reliability.

In addition to the above analytical methods of estimating reliability, the paper examines conditions under which the running time for Monte-Carlo simulation of networks necessary for 'good' estimation of network reliability will grow only linearly with network size. These conditions are:

1. the run time to analyze a single sample network is linear with respect to the network size.

2. the variance of the reliability estimator remains unchanged as network size increases.

It should be noted that methods commonly used to evaluate flows in water supply networks, e. g., the Hardy-Cross method do not have running times that are linear with respect to network size. Rather, these evaluation methods have running times that increase with the square of network size.

The second condition follows from the definition of the variance of an independent sample. If the variance of a random variable X for a single sample is S_{est} then the variance of the mean $\overline{X}$ for n independent samples is S_{est}/n. Thus, if the variance of the random variable remains constant, as network size increases the number of simulation runs necessary to reduce the variance to a given level also remains constant. No results were presented showing how the variance of network reliability changes with network size.

Furthermore, no results are presented under which these conditions would hold true. However, several results are presented which show that series-parallel reductions reduce the variance of reliability estimators.

Combinatorial properties of directed graphs useful in computing network reliability, Satyanarayana and Hagstrom, (1981)

In this article the network reliability problem is again presented in terms of the reachability problem. In this formulation one node is designated as the source and all the other nodes are demand nodes with directed flow from the source to the terminal nodes. The actual method to evaluate network reliability presented in this article is not of interest because it requires the enumeration of all acyclic spanning t-subgraphs of the graph. This task requires on the order of $(n-1)!$ operations, where n is the number of nodes in the network.

However, the basic approach, which is to apply the inclusion-exclusion probability theorem to the calculation of reachability, is of interest. This theorem calculates the probability of a union of n events as the sum of the intersections of the events taken successively $1, 2, \ldots, n$ at a time. The formula to determine this probability is:

$$\begin{aligned} P(\cup A_i) = \sum_{i=1}^{n}(P(A_i)) - \\ \sum_{i,j}(P(A_i \cap A_j)) + \cdots \end{aligned} \tag{2.1}$$

where

- A_i is the event that spanning tree i is working
- $\cup A_i$ is the union of all possible A_i
- $P()$ is the probability that an event occurs

In this form, the inclusion-exclusion theorem is not very useful because it is based on spanning trees. Further articles, reviewed below, consider more useful formulations of this equation.

Network reliability and the factoring theorem, Satyanarayana and Chang (1983)

In this article the concept of the domination of a graph, presented in the previous article (Satyanarayana and Hagstrom, 1981), is extended and combined with a method of factoring a graph, to produce a recursive algorithm for the calculation of network reliability. As this algorithm also requires on the order of $(n-1)!$ calculations, where n is the number of vertices, the algorithm is of little practical significance.

However, once again this article is chiefly of interest because the inclusion-exclusion theorem is presented and reviewed for applicability for both success sets and failure sets. The observation is made that a success set is a minimum spanning tree and that, conversely, a failure set does not contain such a spanning tree. Both formulations are found to require an exponential number of calculations, with respect to the network size to determine reliability. The formulation of the theorem based on the failure sets version, which is the basis of the work in this thesis, is:

$$P(S) = P(S_1) + P(\bar{S}_1 \cap S_2) + \cdots + ((\bar{S}_1 \cap \bar{S}_2 + \cdots + \bar{S}_{n-1}) \cap S_n) \tag{2.2}$$

where S_i is the event that a spanning tree is connected.

While in the general case the use of Eq. 2.2 requires an exponential number of calculations for the calculation of reliability, the formula has a deceptively simple advantage for the estimation of reliability. This advantage is that it orders failure sets by the number of elements in the failure set. As a general rule, all other things being equal, the more system components that must fail for a failure event to occur, the less likely the event is to occur. This general rule is useful in applying Eq. 2.2 to estimate rather than calculate reliability because estimation techniques reduce the number of required calculations by leaving terms out of the above expression. Consequently, it can be expected that a good method for the estimation of reliability is simply to truncate the above formula after a sufficient number of terms, N_{max}, have been calculated.

An impediment to the application of Eq. 2.2 has been the inability to determine the error associated with truncating the series to N_{max} terms. One of the signicant contributions of this thesis is to show that historical break frequency data for urban water distribution systems can be used to limit N_{max} to a reasonable number of terms. In addition, methods to estimate the error introduced by limiting the size of N_{max} are demonstrated in this thesis.

Calculating bounds on reachability and connectedness in stochastic networks, Ball and Provan (1983)

In the article computational procedures are presented for the calculation of bounds on reachability, (defined as the probability of all nodes in a connected network being connected to the source node.) These bounds are presented as polynomials in p where p is the common failure probability for all network links. It is assumed that the link failure probability p is known with certainty. The basic form of polynomial that is used to calculate reliability of a network is:

$$g(G, s; p) = \sum_{j=0}^{n} f_j p^j (1-p)^{n-j} \tag{2.3}$$

where f_j is the number of sets of arcs, of cardinality j, whose complements contain a spanning tree, of the network G, rooted at source node s.

It should be noted that this polynomial in the probability of link failure is similar to the one used by Ball and Nemhauser (1979), where the reachability problem is formulated as a stochastic coherent binary system. The upper and lower bounds for reliability are expressed as polynomials based on f_j. The polynomials are the best general method of calculating network reliability for the case in which links fail independently and have an equal probability of failure. These polynomials are quite complex and use a number of symbols not found elsewhere in the literature. These definitions are necessary to make the polynomials used to bound system reliability understandable. Consider the k-canonical representation of a non-negative integer m as given by:

$$m = C(m_k, k) + C(m_{k-1}, k-1) + \cdots + C(m_1, 1) \tag{2.4}$$

where

- $m_k > m_{k-1} > ... > m_l \geq l \geq 1$
- C(m,k) is the number of combinations of m things taken k at a time.

The m_i are chosen successively, such that

$$m_i = max(x : C(x,i) \leq m - \sum_{j=i+1}^{k} C(m_j, j)) \tag{2.5}$$

The paper claims that the k-canonical representation always exists and is unique.

Before defining the bounding polynomials, the paper defines the upper and lower pseudopowers of an integer sequence. For an integer sequence $(m_k, \ldots, m_l)$, $k \geq l \geq 1$, and any $i \geq k$ define the (i,k)th lower pseudopower of $(m_k, \ldots, m_l)$ to be

$$m^{i/k} = (m_k, \cdots, m_l)^{i/k} = C(m_k, i) + C(m_{k-1}, i-1) + \\ \cdots + C(m_l, i-k+l) \tag{2.6}$$

where i/k denotes the lower pseudo power. Note the role played by the integers i and k in the series in Eq. 2.6.

and the (i,k)th upper pseudopower of $(m_k, \ldots, m_l)$ to be

$$m^{<i/k>} = (m_k, \cdots, m_l)^{<i/k>} = C(m_k - k + i, i) + \\ C(m_{k-1-k+i}, i-1) + \cdots + C(m_l - k + i, i - k + l) \tag{2.7}$$

where

- $< i/k >$ denotes the upper pseudopower. Again note the role played by the integers i and k in the series Eq. 2.7.

The (i,k)th lower and upper pseudopowers, for an integer m, are then defined as the pseudopowers of the k-canonical representation of m. The lower and upper pseudopowers of m are written as $m^{i/k}$ and $m^{<i/k>}$, respectively.

Polynomial expressions in p are then presented for the upper and lower bounds on network reliability. The formula for the lower bound $gL(G,s;p)$ is:

$$gL(G,s;p) = \sum_{i=0}^{k} f_i p^i (1-p)^{n-i} + \sum_{i=k+1}^{d} f_d^{i/d} p^i (1-p)^{n-i} \tag{2.8}$$

The formula for the upper bound on network reliability gU(G,s;p) is:

$$gU(G,s;p) = \sum_{i=0}^{k} f_i p^i (1-p)^{n-i} + \sum_{i=k+1}^{d-1} f_k^{i/k} p^i (1-p)^{n-i} + f_d^{i/d} p^d (1-p)^{n-d} \tag{2.9}$$

where

- f_i, G, s, p are as defined previously
- $d = n - q + 1$,
- n = number of links in the network
- q = number of nodes in the network

It is shown that for important classes of graphs it is possible to calculate bounds on reachability in polynomial time. These classes of graphs include acyclic directed graphs and planar graphs. The case of planar graphs is of particular interest to this study because water supply networks can be successfully modeled as planar networks.

It can be seen that the width of the calculated bounds on reliability are dependent on the failure probability p. For values of p that are less than 0.01 the bounds are quite narrow, on the order of .01. For higher failure probabilities, greater than 3-4%, the width of the bounds can be as great as 0.2. This large range in the reliability bounds limits the practical usefulness of this method for networks with high component failure probabilities. In addition, the assumptions that the probability of failure is known and equal for all links is highly unrealistic. As a result, this article is chiefly of interest because it gives an indication of the trade-off necessary to determine the probability of network failure exactly versus the use of estimation techniques to approximate it. The paper is encouraging in that it indicates that it is possible to approximate, even if roughly, reliability in a polynomial number of operations for the class of planar graphs.

A survey of network reliability and domination theory, Agrawal and Barlow (1984)

This article surveys the computational complexity of various methods of calculating network reliability and applies domination theory to series-parallel reducible graphs. The authors note that in the general case this problem is NP-hard. NP-hard problems are distinguished in this article as being harder than NP-complete problems. An NP-complete problem with n basic elements, e.g., n nodes or links, has the properties that

1. the number of operations to solve it is greater than the number of operations in any polynomial in n

2. the solution can be checked in a polynomial number of operations.

The computation time to solve NP-complete problems increases very rapidly with problem size. An NP-hard problem conforms to property 1, but has the additional feature that the solution (reliability) cannot be checked without solving it. The paper points out that the network reliability problem is NP-hard. This observation is important from a modelling point of view because most reliability based optimization design techniques for water distribution network design require that the change in reliability with respect to a decision variable be determined. Previous work, e.g. Su et al., (1987) approximated the change in reliability with respect to network cost by evaluating network reliability many times. As calculating network reliability for NP-hard problems even once is computationally impracticable, calculating it many times at each iteration of the design process to determine the gradient, is essentially impossible.

The solution methods discussed in the paper are the inclusion-exclusion method, the sum of disjoint products, and factoring. In light of the previous discussion, all of the methods presented can be expected to require excessive numbers of calculations in the general case. The paper does show that pivotal decomposition can be used to calculate reliability for series-parallel reducible graphs in polynomial time. This class of graphs is slightly more general than previously known classes of graphs, for which reliability can be calculated in only polynomial time. The class of graphs is not, however, generally applicable to water supply problems because any network that has three or more links intersecting at any interior node cannot be completely reduced.

Evaluation of methods for decomposition of water distribution networks for reliability analysis, Jacobs and Goulter (1988)

This article reviews previous work on the design of reliable water distribution networks with a view toward the application of graph theory to the specific problem of water distribution networks. Review of the methods of directly calculating network reliability resulted in the conclusion that all such methods are impractical because of excessive computational requirements.

Methods of reducing calculations, based on empirically derived filtering techniques, were also reviewed. In particular, the method of N-k cuts was discussed. In this method, it is assumed that cut sets are limited to a size less than or equal to k. A number of examples were presented to demonstrate the inadequacy of this method, when k is chosen arbitrarily. Other methods based on filtering techniques examine the use of total head loss and head loss gradient, to eliminate paths from consideration in reliability calculations. It was found that these methods did not reduce the required number of calculations sufficiently to make the problem computationally feasible.

In addition, a number of heuristic methods for the calculation of network reliability were discussed. These heuristics included node based measures that require a specific number of edges to be incident on each node to prevent the isolation of the node. The main disadvantage identified for these measures was that, even for small networks, the heuristics tend to overestimate network reliability. No heuristic methods that took the relative importance of links or nodes in the network into account were identified.

Finally, a theoretically optimally reliable network configuration for undirected graphs with links having equal cost and reliability was proposed. It was hypothesized that this optimally reliable configuration could provide a model or target (upper bound) for an optimally reliable water supply network, without the necessity of having to calculate network reliability explicitly.

Optimization of redundancy in water distribution networks using graph theoretic principles, Jacobs and Goulter (1989)

In this article the hypothesis that the optimally reliable network configuration concept introduced in Jacobs and Goulter(1988) could be used to design optimally reliable water supply networks was tested using an integer goal programming model. The model attempted to make the degree of each node equal, a condition necessary for a network to have an optimally reliable configuration. Manual intervention was required in the model to eliminate weaknesses such as disconnections, bridges, articulation nodes, and implicit bridges and articulation nodes that this definition of regularity did not preclude.

It was possible to design networks that looked reasonable using this model by making several runs to eliminate the above weaknesses. These designs had the advantage that redundancy was an integral part of the design.

One weakness of the model was that it does not differentiate between areas of the network. All areas of the designs have the same relative density of links. Efforts to get the model to differentiate between nodes by weighting them were not successful. It was concluded that the importance of a node to the overall network is in large part due to demands of downstream nodes rather than demand at the node. This conclusion suggests that the assumption that a water supply network can be adequately modeled as an undirected graph is invalid. This is not a particularly restrictive issue as the direction of flow in each pipe can be uniquely determined for a given set of pipe sizes and nodal demands. However, it does imply that water supply networks must be modeled as directed networks. However, because the flow in many of the pipes in such a network is reversible, under small perturbations in flow or hydraulic head, provision for flow in both directions is necessary.

Problems were also encountered in terms of the computational effort required for this model. One run failed after 25,000 iterations of the branch and bound technique used to solve the integer program. This difficulty suggests that, in the worst case, the computational requirements for this method are excessive.

Estimation of maximum cutset size for water network failure, Jacobs and Goulter (1991)

In the article a method to estimate network reliability, using the cut set method is developed. It has been shown that the calculation of network reliability is NP-hard using cut sets. Current engineering practice, however, uses cut sets of size one (Endrenyi, 1978). The basis of this practice is that cut sets containing multiple elements are improbable.

The approach involved the calculation of network reliability is decomposed into the determination of the probability of a given number (k) of simultaneous pipe failures and the probability of network failure, given that k pipes have failed. The probability of k simultaneous failures, used in the demonstration of the procedure, was estimated from the City of Winnipeg pipe break data base, using 10 years of data. In assessing the likelihood of simultaneous failures the procedure assumes that all breaks that occurred on the same calendar day were in fact simultaneous. It was also assumed that all links had the same probability of failure and were equally reliable.The probability of failure, given k simultaneous breaks, was estimated for the network of Morgan and Goulter (1985) using a simulation program.

Regression was performed on the distribution of k=1, 2, 3 etc. breaks for the City of Winnipeg to estimate the probability of that number of breaks occurring for a network the same size as the one presented in Morgan and Goulter (1985). It should be noted that this method of scaling down from the City of Winnipeg's network, to the one presented in Morgan and Goulter (1985), assumes that breaks are evenly distributed throughout the network. Since failures are known to be clustered in time and space, further study is required to scale k more accurately with system size.

The methodology presented in the paper appears to be quite promising as a method of limiting the required calculations, for the determination of network reliability. However, many limiting assumptions were made and these require further development.

Spatial and temporal groupings of water main pipe breakage in Winnipeg Goulter and Kazemi (1988)

The results shown in Jacobs and Goulter (1991) demonstrate that the pipe failure rate and pattern specific to a given location can be useful information in determining the reliability of pipe network located at that location. In the article, pipe breakage, specifically, the dependence of the occurrence of pipe breaks on previous pipe breaks, in the City of Winnipeg, Canada is investigated. A computerized data base containing break records for the City of Winnipeg from 1975 through 1985 was used in the analysis. The occurrence of breaks at varying distances and durations from a previous break are examined. The distances are varied from 0 to 20 meters. The durations are varied from 0 to 90 days.

The results of the analysis showed that a strong clustering effect exists for pipe breaks in the City of Winnipeg. The authors found that 22% of all breaks occur within one meter of another break and 46% of all breaks occur within 20 meters of another break. In addition, 9% of all breaks occur within 1 day and 1 m of another break and 42% of all breaks that occur in within 1 m of another break also occur within 1 day of the initial break.

Shifts in the soil surrounding a break due to water leakage in the vicinity of a pipe break was suggested as being a cause of the clustering effect. Soil disturbance necessary to repair a pipe break was also suggested as being a contributing cause to the subsequent breaks. It was also noted that the pipe breakage rate is higher in the winter than the summer indicating that exposure of the pipes to extreme cold temperatures during the winter and early spring might be a cause of subsequent breaks. (It is common for winter temperatures in Winnipeg to be in the -20 to -40 degree Celsius range).

No experimental evidence is presented in this paper to substantiate these suggestions. It should also be noted, that while large variations in breakage rates are cited in this article for pipes of different diameters and made from different materials, no separation of breakage rates by pipe material and size is made. One of the contributions of this thesis is to break down failure patterns by pipe diameter, pipe size, and pipe age. Another weakness of the paper is that no reference is made to the physical process by which breaks occur, i. e., rusting for cast iron pipes.

Despite these weaknesses, the evidence for the existence of the clustering pattern described in the paper is quite strong and provides the basis for the modified Poisson model presented in Chapter 6.

2.3 Estimating reliability: bounding the error

Combining Monte Carlo estimates and bounds for network reliability Nel and Colbourn (1990)

In this article Monte Carlo simulation is used to estimate the coefficients for the reliability polynomial presented in Ball and Provan (1983). The reliability calculation is based on the assumption of a known and equal probability of failure for all links. In addition, it is assumed that failures are independent events. It should be noted that in practice the probability of failure of a given link is not known. Similarly, the assumptions of independence and equiprobable failure are unrealistic for urban water supply networks. A more complete discussion of the independence issue is given in Chapter 6. As a result of these limitations, the results presented in this article are not directly applicable to the urban water supply problem.

This article is a valuable contribution, however, because it recognizes the trade-off between estimating reliability and the extreme computational burden necessary for the exact calculation of reliability. The article also addresses a number of steps that must followed in using Monte Carlo simulation. These steps are identified as: determining the sample space, determining required sample size, dividing the sample space into discrete sub-intervals, and estimating simultaneous confidence limits for the discrete sub-intervals.

In addition to the problems mentioned previously concerning the applicability of this method to water supply distribution networks, the method by which the elements of the sample space are identified is highly dependent on the assumptions of independence and known equiprobable failure.

The sample events are identified as the set of spanning trees. A key part of the article is the development of an algorithm to separate a graph into spanning trees. Ball and Provan (1983) showed that all spanning trees have the same number of links. This result follows since all spanning trees must have the one link less than the number of nodes in the network. Since Ball and Provan (1983) assumed all links have the same probability of failure it follows that under this assumption all spanning trees must have the same number probability of failure. Therefore, in the case of a network in which all links have an equal probability of failure, the spanning trees subdivide the graph into equiprobable subsets. The spanning trees can then be used as a basic sampling unit in the Monte Carlo simulation. It is difficult to see how this approach could be generalized to a real network comprised of intermixed pipes of different sizes, materials, and ages. Of necessity, the spanning trees would contain pipe which differed in diameter, construction material, and age. As a result the probability of failure of the different spanning trees would be different in spite of the fact that they contain the same number of links. These problems with this method make it extremely difficult to generalize to more realistic networks.

However the difficulties found in generalizing the results in the article, highlight the importance of using an approach to the calculation of reliability which can easily be generalized to reflect conditions in more realistic networks. In particular, these real networks have components which do not fail independently and which do not fail with equal probabilities. It might, in fact, be preferable to use a method of estimating reliability which is less accurate, but which accurately reflects the properties of network components, than one which is very accurate but cannot be generalized.

A comparison of four Monte Carlo Methods for estimating the probabilities of s-t connectedness, Fishman (1986)

In this article four Monte Carlo methods are presented for estimating the s-t (source to terminal) reliability problem. The s-t reliability problem evaluates the ability of the network to serve a single demand node. Since urban water distribution networks are generally evaluated for reliability on the basis of whether or not all node are served adequately the s-t reliability problem is not generally applicable to estimating water supply distribution network reliability. As the reliability problem solved in the paper is not directly applicable to the evaluation of urban water distribution networks, the detailed evaluations of the various methods of Monte Carlo simulation are not applicable to this study.

This article is of interest chiefly because it contains a good basic explanation of the Monte Carlo process. The paper asserts that the Monte Carlo sampling process follows a binomial distribution. An estimator of the probability that the network does not fail, g, is given as:

$$\overline{h_K} = \frac{1}{K}[\sum_{k=1}^{K} \phi(x^{(k)})] \tag{2.10}$$

where

- K = the number of trials

- $\phi(x^{(k)}) = 1$ if network does not fail and 0 otherwise

The variance of h_K is then given as:

$$Var(h_K) = \frac{g(1-g)}{K} \tag{2.11}$$

This particular formulation for the variance of a Monte Carlo simulation package is used in this thesis.

Modelling and analysis of systems with multimode components and dependent failures Le and Li (1989)

In this article the authors describe a method of estimating the reliability of a network with components that can operate at several performance levels and in which failures are not independent. Previously reviewed articles have assumed that failure events are statistically independent. The paper claims that the assumption of statistical independence is not realistic because neighbouring components are likely to be exposed to similar environmental stresses and because several components may depend on the same common physical resource.

Examples of both of these effects are given for communications networks. It is not, however, difficult to find ways in which both modes of dependence can apply to water distribution networks. For example, if a section of pipe is located in an area where the soil is corrosive or there is a high electrical grounding load, high rates of rusting in cast iron pipe can be expected. The higher rusting rates can be expected to contribute to higher failure rates. It can also be expected that neighbouring sections of pipe are exposed to similar corrosive soils and/or electrical grounding loads. In addition, if two demand nodes receive water supply from the same source both are affected if the flow from the source is interrupted. These examples suggest that the problem examined in the article deals with a realistic problem.

The approach used to solve the problem is to classify failures by their causes and assumes that the effects of each cause are measurable, separable, and independent. These conditions are not fulfilled for water supply systems. Highly corrosive soils and increased stress caused by soil movements are probably two causes of pipe failures. The highly corrosive soils can cause small leaks which do not significantly impair the functioning of the pipe. These small leaks, however, change the moisture balance in the soil, which can cause the soil to shift. The shifting soil can then cause the pipe to fail. Since the assumptions used for the method do not fit water distribution systems, the method is not directly applicable to calculating the reliability of water distribution systems.

However, the calculation method used in the article is similar to the approach taken in this thesis. As a result, several issues raised in the article are of interest. The first issue is the basic form of the expression used to calculate reliability. This expression is given as:

$$R = \sum_{s_h \in \Omega} P(s_h) R(s_h) \tag{2.12}$$

where

- Ω is the set of system states
- s_h is the hth system state
- $P(s_h)$ is the probability of s_h occurring
- $R(s_h)$ is the reliability measure of the system given that s_h occurs

In general, to calculate reliability exactly, the number of terms in this expression grows exponentially with system size. As a result, the approach of truncating Eq. 2.12 was proposed. The system states are labelled in order of decreasing probability and sufficient terms are included in the truncated version of Eq. 2.12 to ensure that the resulting error is acceptable.

This approach is similar to the one that is presented in Chapter 5 of this thesis. Several problems, with the approach as described in the paper should, however, be recognised. The first is the assumption that when the events s_h are labelled so that their failure probability decreases monotonically, the reliability measure increases monotonically. However, there is no a priori reason to assume that statistically unlikely events are likely to cause system failure or that highly probable events are unlikely to cause system failure. 'Configuration' of the events as it relates to graph topology impacts on the likelihood of system, as opposed to individual component, failure. A second problem is that the paper does not take into account the uncertainty resulting from the estimation of $P(s_h)$ and $R(s_h)$. This uncertainty adds to the difficulty of applying this approach.

Despite the limitations mentioned, the approach to estimating reliability described in the article, has the advantages of being able to solve a realistic problem in a robust fashion.

2.4 Solving a Hydraulic Network

Efficient code for steady-state flows in networks, Epp and Fowler (1970)

The article describes a computer code to set up and solve a hydraulic network using the Newton-Raphson method. The Newton-Raphson method has the disadvantage that the initial estimates of flows in each pipe in the network should be close to the actual or final flows. The paper suggests that the flows in a spanning tree of the network are a good approximation to the actual flows. Unfortunately, the flows in a spanning tree are not necessarily good approximations to actual flows if a significant inflow volume originates in overhead storage tanks rather than source pumps. In addition, setting up the loop equations requires several steps which require extensive programming effort. These steps include an algorithm to determine natural loops and to ensure that a near minimum band-width matrix is created when the loop equations are set up.

For these reasons, the algorithm described in this paper was not used directly in this thesis. However the paper also presents an algorithm to determine the shortest path between two points in a network. This algorithm is used to create loop equations in the simulation program described in Chapter 9.

Hydraulic network analysis using linear theory, Wood and Charles (1972)

In this article a method to solve a hydraulic network based on linear theory is described. The key to the method is the simple way the equations for the head loss around a loop are linearized. More specifically each head loss term is written as:

$$h_{Li} = K_i Q_{i0}^{a-1} Q_i = K_i' Q_i \tag{2.13}$$

Where

- h_{Li} = head loss in pipe i
- K_i = head loss constant
- K_i' = head loss constant to be used as a coefficient in the linear equations
- Q_{i0} = estimated flow in pipe i
- a = head loss exponent
- Q_i = actual flow in pipe i (unknown)

Based on this approximation, a linear set of equations can be written to solve a hydraulic network. Any solution method for a set of linear equations can then be used to solve the network for new flow values, Q_{i0}. The new flow values can then be used to estimate new values for the constants, K_i'. The new constants can then be used to solve the network for new flow values. The process continues until the change between iterations is less than a minimum tolerance for all flow values.

The advantages of the method are that convergence is insensitive to the choice of initial link flow values and that convergence is rapid. The only disadvantage of the method is that it requires more computer memory than the Newton-Raphson method. This disadvantage is minor for all practical sized networks because of the rapid increase in available computer memory since the method was developed.

The authors do not cover methods for the inclusion of storage tanks and pumps in the article. Despite this limitation, the method presented, in this article, forms the basis for the network solver used in the simulation program developed for this thesis.

Pumps and reservoirs in networks by linear theory, Jeppson and Tavallaee (1975)

The article extends the work of Wood and Charles (1972) to include pumps and reservoirs. Reservoirs are handled by creating 'pseudo-loop' equations that set the head loss between two reservoirs equal to the total head loss in pipes that connect them. The method is incorporated in the simulation program created for this thesis.

Pressure reducing valves in pipe network analysis, Jeppson and Davis (1976)

The article provides a good summary of the linear theory method for solving a hydraulic network. It also extends the linear theory method to networks with pressure reducing valves. The article was primarily used in this thesis because of the clear explanation of the method used to develop equations to include reservoirs in the hydraulic network.

Chapter 3

Difficulties in the Exact Calculation of Network Reliability

Several references in the literature have shown on a theoretical basis that the calculation of reliability is computationally intensive. In particular, Ball (1980) proved that the calculation of network reliability is in the class of NP-hard, under the simplifying assumptions that:

1. failure is equivalent to node disconnection
2. links have equal failure probabilities
3. failure events are independent

The proof shows that the number of steps to calculate reliability increases exponentially with system size. The only way in which an algorithm, for the exact calculation of reliability in which computational requirements do not increase exponentially with system size, could be found is if a number of other computationally difficult problems could also be solved in a number of steps that is bounded by a polynomial function of the problem size. These other computationally difficult problems include the travelling salesman problem and the satisfiability problem in combinatorics. No polynomial algorithm for these problems has been found, in spite of more than 15 years of intensive effort.

A cursory examination shows that the number of calculations can become very large for even moderately sized networks. For a network with 17 links, approximately 130,000 calculations are required; while for a network with 30 links, approximately 10^9 calculations are required.

An issue that should be considered, however, is the possibility that, while the computational requirements to determine system reliability are excessive in the worst case, these requirements may be reasonable for most networks found in practice. One way of approaching this question is to consider the types of networks for which the computational burden to determine reliability is limited to a reasonable effort. The two types of networks for which the computational burden is limited to a reasonable level are trees and series-parallel reducible networks.

A typical example of a tree network is shown in Fig. 3.1. The calculation of network reliability for a tree is straightforward because of the topology of the network. Specifically, a tree has a unique path, connecting the source node with each node of the network. For a network with n nodes, reliability can be calculated by calculating the probability of failure of the path to each node. The network reliability is then given by:

$$R = 1 - \sum_{i=1}^{n} P(F_i) \tag{3.1}$$

where

- R = system reliability
- $P(F_i)$ = probability that the path to the ith node fails

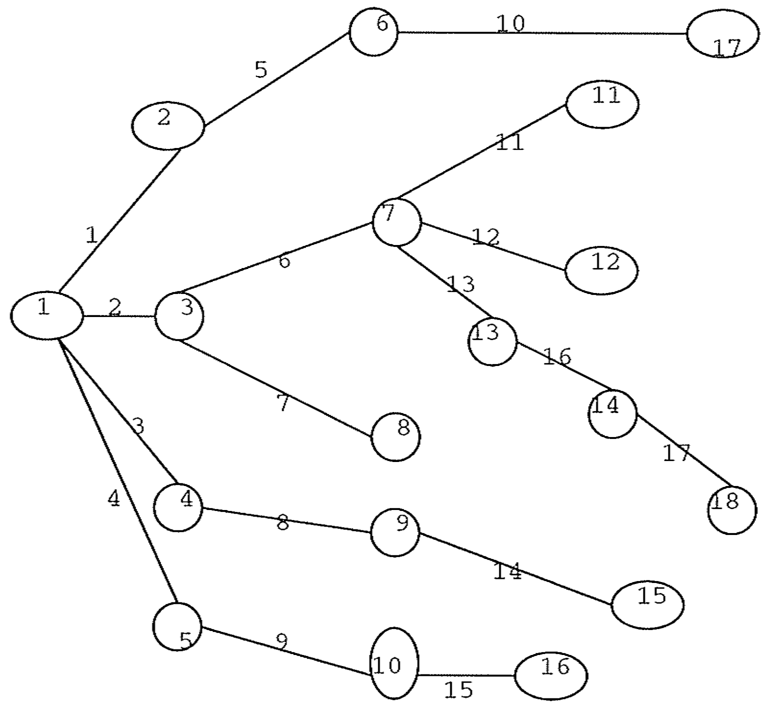


Figure 3.1: Example of a Tree Network

The same network topology, however, makes a tree inherently unreliable. If a network has only one path, connecting a node to the source node, then the failure of any link in the network causes at least one node to be disconnected, i.e., network failure occurs under both connectivity and hydraulic criteria. In addition, the impact of the network failure increases for the failure of links that are closer to the source node. For example, in Fig. 3.1, the removal of link 2 would disconnect approximately one half of the network.

The situation, with respect to series-parallel reducible networks is somewhat more complicated. The calculation of reliability, with a reasonable computational effort, for a series- parallel reducible network involves transformation of the network topology. Blocks of the network are replaced with single complex links. These links must contain sufficient information to characterize the reliability characteristics of the blocks they replace. If the network is series-parallel reducible, the transformed network is reduced to a tree. The mechanics of this process are described in Agrawal and Barlow (1984) who also show that all graphs are not series-parallel reducible.

Fig. 3.2 shows a typical series parallel reducible graph. Again, the topology of the network is instrumental in making the reliability calculation tractable. The presence of interior nodes, i.e., nodes not connected directly to either the source (s) or terminal (t) nodes, is a key determinant of whether the network is series-parallel reducible. It should be noted that the network in Fig. 3.2 has no interior nodes of degree three or greater. Any network that has more than three nodes, of at least degree three, is not series-parallel reducible. After all series-parallel reductions have been performed for such a network a series-parallel reduced network remains. The computational effort required to calculate reliability for this reduced network depends on its size. If the series-parallel reductions have reduced the size of the original network to a small size, approximately, 10 links or less, then the calculation of reliability is feasible. On the other hand if the reduced network contains many more than 10 links then the computational effort to calculate reliability remains excessive despite the series-parallel reductions.

The topology necessary for a network to be series-parallel reducible can have adverse effects on the network's reliability. In particular, efficient design of water supply distribution networks requires that pipe capacity be reduced with distance from the source node. The increased capacity, in links near the source node, is necessary because these links must carry water to supply both the node at the end of the link and demand nodes, throughout the network. The links that are far from the source node are designed to carry water to supply only the demand nodes in the neighbourhood of the links.

The implication of this reduction in capacity is that it is impossible to meet demand requirements for nodes, located far from the source, by reversing the normal

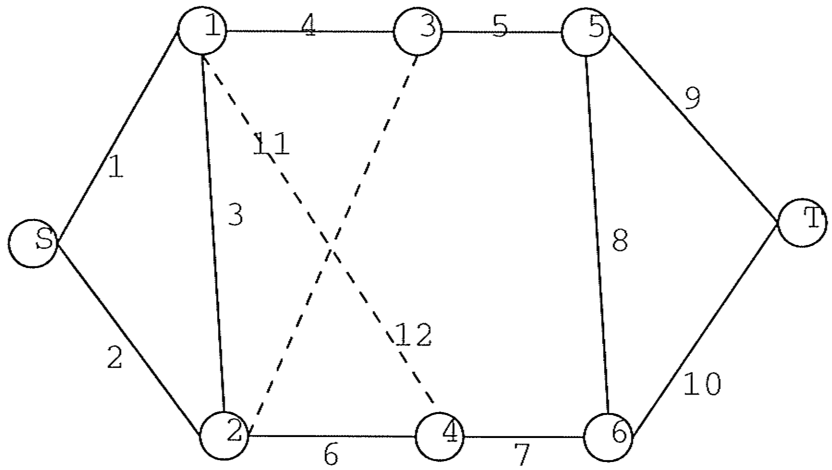


Figure 3.2: Example of a Series Parallel Reducible network

direction of flow in the adjacent links. If, in Fig. 3.2, link 4 fails, node 3 is still connected, but it is quite possible that the connecting pipes do not have sufficient capacity to fully supply node 3. The addition of links 11 and 12, shown as dashed lines in Fig. 3.3, would clearly improve the reliability of the network. It should be noted, however, that the addition of these links would change the topology of the network, so that it is no longer series-parallel reducible.

3.1 Comments

The previous examples suggest that although not as forbidding as general theoretical analysis would suggest the problem for realistic networks is still too large to be calculated with reasonable effort. Conversely, it is often the case for large networks that, if the network is designed so that reliability can be calculated with a reasonable effort then the network does not have sufficient reliability. Furthermore, the above examples assumed highly simplified failure criteria and failure mechanisms. However, water supply systems can fail by providing insufficient flow or pressure while all nodes are connected. In addition, experience shows that pipe failures are not independent that that failure likelihoods are not uniform throughout the network.

The above problems with the exact calculation of network reliability justify the exploration of sampling methods to approximate reliability rather than calculating it exactly. In this way that the computational burden associated with calculation of reliability can be reduced to reasonable levels for quite large networks. The use of estimation techniques is especially justified if as will be shown the estimation methods can be used to relax the assumptions concerning failure mechanisms and likelihoods. The opportunity to relax these assumptions would make the reliability estimates far more useful.

3.2 Use of Conditional Probability

The basis of the method used to approximate network reliability is the general multiplication rule of the theory of probability. This rule states the probability that a pipe failure event causes a network failure as follows:

$$P(E \cap F) = P(E)P(F \mid E) \tag{3.2}$$

where

- F is the event that the network fails
- E is the event that a set of pipes in the network fails
- $P()$ is the probability function

This formulation is particularly useful because it divides the network failure event into a term $(P(E))$ which is entirely dependent on the component reliability of the network and a term $(P(F \mid E))$ which depends entirely on the network topology. The first term, $P(E)$, can be estimated from failure history of the pipes in the network in question. The second term, $P(F \mid E)$ can be estimated using Monte Carlo simulation. In applying this decomposition the component failure event E is divided into disjoint sets that involve the simultaneous failure of exactly m = 1, 2, 3, etc. links in the network. These disjoint events are mutually exclusive and collectively exhaustive of the component failure space. If the component failure event is divided into m, mutually exclusive, exhaustive events E_1, E_2, E_3, etc. the equation for network failure, Eq. 3.2, becomes:

$$P(E \cap F) = P(E_1)P(F \mid E_1) + P(E_2)P(F \mid E_2) + P(E_3)P(F \mid E_3) + \ldots \quad (3.3)$$

where

- F is the event that the network fails
- E_k is the event that exactly k pipes in the network fail
- $P()$ is the probability function

Eq. 3.3 provides the basis for the approximation of network failure probability used in this study. In Chapter 4 a model is developed which can be used to estimate system failure probability under the assumption that pipes in the network have the same probability of failure.

Chapter 4

A Simplified Model for Bounding Failure Probabilities

In Chapter 3, the limiting assumptions and calculation burden associated with current work on network reliability was discussed. Since, at the present time, there is no model which adequately addresses these issues, a mathematical model which bounds network reliability is presented. The model is first developed in this chapter for the simplified case in which all links have an equal probability of failure and in which failure events are independent. In Chapter 5, the model is extended by relaxing these simplifying assumptions.

4.1 Qualitative Basis of the Model

Two factors affect network reliability when components can fail. The first factor is the failure probability for one or more individual components. In particular, when effects on service levels are considered, the event that several components fail simultaneously is far more serious than the event that the same number of components fail one at a time. As a result, information about the simultaneous failure of multiple system components can be expected to be important in the development of a reliability estimation model.

The second factor that affects network reliability is the geographical location(s) of a component failure or combination of component failures. A small number of failures, 'correctly' positioned, can cause a system to fail. On the other hand, a greater number of randomly placed failures might not cause system failure if sufficient redundancy is built into it. The geographical locations of component failures in a pipe network are subject to stochastic variation and, hence, can not be predicted exactly.

To quantify the combined effects of these two factors the probability of failure is decomposed into a simultaneous component failure term and a 'geographical' reliability term. The simultaneous component term can be estimated using historical data for the network being studied. The geographical reliability term can be estimated using Monte Carlo simulation. In the next section, a mathematical model based on these principles is developed.

4.2 Theory

The mathematical model to be developed uses graph theory to represent a water supply distribution network. A network, with ML pipes can be represented as a graph with ML links. Under the simplified assumptions being used in this chapter, it is assumed that links fail independently, have an equal probability of failure, and the network is considered to have failed if one or more nodes have been disconnected from the network. Denote the event that the network fails by F.

Since water distribution networks are designed with some redundancy, the failure of one or more links does not necessarily lead to network failure. A set of failed links, which disconnects one or more nodes from all source nodes, is called a cut set. Any set of links that fails is called a removal set. Denote the removal set consisting of at most m failed links as S_m. Note that, implicit in this definition is the requirement that $m > 0$. Denote the event that no links fail as SNF. The definition of S_m and the fact that m is an integer implies that the set S_m contains the events that $1, 2, 3, \ldots or\ m$ links are removed from the network. S_m does not necessarily cause the network to fail. The following property holds for S_m

$$S_1 \subset S_2 \subset S_3 \ldots \subset S_{ML} \tag{4.1}$$

where

- S_m is the event that $1, 2, \ldots, or\ m$ of the links are removed from the network
- ML is the number of links in the network

The following union covers the set of all possible removal sets:

$$SNF \cup S_{ML} \tag{4.2}$$

where

- SNF is the event that no links are removed from the network
- S_{ML} is the event that $1, 2, \ldots, or\ ML$ links are removed from the network

It is assumed that failure of the network is caused by failure of links in the network. Thus, the event that the network fails and at most, m links are removed, is denoted by $(F \cap S_m)$. From Eq. 4.1

$$(F \cap S_1) \subset (F \cap S_2) \subset (F \cap S_3) \ldots (F \cap S_{ML}) \tag{4.3}$$

where

- F is the event that the network fails

Because the intersection of S_m with any another set can contain no more elements than were originally in S_m, it follows that

$$(F \cap S_m) \subset S_m$$
$$and\ P(F \cap S_m) \leq P(S_m) \tag{4.4}$$

where

- F is the event that the network fails
- $P()$ is the probability of the event enclosed in the brackets

For water distribution networks, it can be argued that there exists a value N_{max}, which is considerably smaller than ML, for which Eq. 4.2, with N_{max} substituted forML almost covers the removal space. Combining Eqs. 4.2 and 4.4

$$P(SNF \cup S_{ML}) > P(SNF \cup (F \cap S_{ML})) \tag{4.5}$$

Subtracting $P(SNF \cup (F \cap S_{Nmax}))$ from both sides of Eq. 4.5 gives:

$$P(SNF \cup S_{ML}) - P(SNF \cup (F \cap S_{Nmax})) >$$
$$P(SNF \cup (F \cap S_{ML})) - P(SNF \cup (F \cap S_{Nmax})) \tag{4.6}$$

Note that $P(SNF \cup S_{ML}) = 1$ and the left hand side of the equality in Eq. 4.6 is the error introduced by considering removal sets of size N_{max} or less. Eq. 4.6, therefore, shows that $1 - P(SNF \cup (F \cap S_{Nmax}))$ forms an upper bound for the error, in estimating network failure, introduced by choosing a removal set with less than ML links. The error in the probability of network failure caused by the omission of larger cut sets is can thus be bounded using the failure probabilities for the chosen smaller removal sets.

While the above result is useful, the events S_m are difficult to work with because they are not disjoint. The overlapping of the S_m prevents the easy calculation of the probabilities of combinations of events. The disadvantage of formulating the model, in terms of S_m, is especially evident if the model is to be generalized to the case discussed in Chapter 5 where failure probabilities are not equal.

A partitioning of the removal set space divides the space into disjoint sets. Under a partitioning all possible removal sets are in one, and only one, of these disjoint sets. There may be several ways of partitioning the removal set space. A straight-forward and intuitive method of partitioning the removal set space is by the cardinality of the removal events. Under this partitioning all removal events in which the same number of links fail are grouped into a single set. More formally, denote the set of events in which exactly m links fail as E_m, $m = 0, 1, 2, 3, \ldots, ML$. Further denote the event that the network fails and exactly m links fail as F_m. Note that $F_m = (F \cap E_m)$. The failure behaviour of the network with respect to the sets E_m can be related to the failure behaviour, with respect to the events S_m.

Consider the event in which no links are removed ($m = 0$). This event can be denoted by either E_0 or SNL. The probability of network failure, and either of these events, is given by:

$$P(F_0) = P(F \cap E_0) = P(F \cap SNL) = 0 \tag{4.7}$$

where

- F_0 is the event that the system fails when no links are removed
- E_0 is the event that exactly zero links are removed

Eq. 4.7 shows that the probability of network failure is zero if all links are working. Now consider the case in which $m = 1$. Because the S_m were defined to exclude removal sets with zero links removed, $E_1 = S_1$. Therefore, the probability of failure, for $m = 1$, is given by:

$$P(F_1) = P(F \cap E_1) = P(F \cap S_1) - P(F \cap SNL) = P(F \cap S_1) \tag{4.8}$$

For the case, in which $m = 2$, note that the event that exactly two links fail is the same as the event that at most two links fail, but the event that only one link fails does not occur. The failure probability for $m = 2$ is therefore given by:

$$P(F_2) = P(F \cap S_2) - P(F \cap S_1) \tag{4.9}$$

Similarly for $m = 3$

$$P(F_3) = P(F \cap S_3) - P[(F \cap S_2) \cup (F \cap S_1)] \tag{4.10}$$

But since $S_1 \subset S_2$ Eq. 4.9 becomes

$$P(F_3) = P(F \cap S_3) - P(F \cap S_2) \tag{4.11}$$

Using the same process, for an arbitrary m:

$$P(F_m) = P(F \cap S_m) - P(F \cap S_{m-1}) \tag{4.12}$$

By summing Eq. 4.12 over all values of m and noting that the right-hand side forms a telescoping series, the following expression for calculating reliability can be found.

$$P(F) = \sum_{m=1}^{ML} P(F_m) = P(F \cap S_{ML}) \tag{4.13}$$

If the series in Eq. 4.13 is truncated after N_{max} terms an error is introduced into the calculation of network reliability. It was shown in Eq. 4.6 that a bound for this error can be determined using removal sets of size N_{max} or less. Eq. 4.13 has the advantage that the maximum size of cut set for which the probability of failure must be evaluated in obtaining this bound is N_{max}. However, the computational burden to evaluate the probability of network failure using cut sets of size N_{max} or less can be prohibitive when N_{max} is large. As a result, estimation of network reliability by sampling is still preferred. Note that Eq. 4.13 can be expressed in a form in which it is easier to estimate the $P(F_m)$ terms. By definition $F_m = F \cap E_m$. The following equation results when $F \cap E_m$ is substituted for F_m in Eq. 4.13.

$$P(F) = \sum_{m=1}^{ML} P(F \cap E_m) \tag{4.14}$$

By the definition of conditional probability

$$P(F \cap E_m) = P(E_m)P(F \mid E_m) \tag{4.15}$$

Substituting the right hand side of Eq. 4.15 into Eq. 4.14 yields

$$P(F) = \sum_{m=1}^{ML} P(E_m)P(F \mid E_m) \tag{4.16}$$

It should be noted, that the right hand side of Eq. 4.16 consists of two terms, one which is dependent on link reliability alone and the other which is dependent on the layout, or geography, of the network.

The factor $P(E_m)$ is dependent only on link reliabilities. It can be estimated from break or failure history data for the city in question and, as such, is empirical. The method used to estimate $P(E_m)$ from City of Winnipeg pipe break data is discussed in Chapter 6.

The $P(F \mid E_m)$ factor is dependent on network geometry and represents the probability that the network fails given that m links have been removed. $P(F \mid E_m)$ can be calculated as:

$$P(F \mid E_m) = \frac{K_m}{T_m} \tag{4.17}$$

where

- K_m is the number of removal sets of size m that cause the network to fail

- T_m is the the number of removal sets of size m.

However, while it is possible to calculate both K_m and T_m directly, the computational burden can be excessive. Jacobs and Goulter (1990) found that, in the general case, it would be necessary to evaluate over 10^9 removal sets of size 4 to determine K_m for a network with 100 links. A Monte Carlo simulation approach therefore is proposed. This simulation approach is used to determine $\hat{p}$, an estimator of $P(F \mid E_m)$. During each trial of the simulation process, m links are removed randomly from the network and the network is tested for connectivity. The system is deemed to have failed if any node is disconnected. Define i to be the number of trials in which the removal of m links results in network disconnection. Denote the total number of trials as t. The estimator of $P(F \mid E_m)$ is thus

$$\hat{p}_m = \frac{i}{t} \approx \frac{K_m}{T_m} = P(F \mid E_m) \tag{4.18}$$

where

- $\hat{p}_m$ = estimator of $P(F \mid E_m)$

- i = number of trials that cause the network to fail

- t = total number of trials

If all possible removal combinations are tested, an estimate of the probability of system failure becomes

$$P(F) = \sum_{m=1}^{ML} P(E_m)\hat{p}_m \tag{4.19}$$

In the development of Eq. 4.6 it was hypothesized that there exists a value, $N_{max} < ML$, for which $S_{N max}$ includes most removal sets. If N_{max} is used in place of

ML, Eq. 4.19 becomes

$$P(F) = \sum_{m=1}^{N_{max}} P(E_m)\hat{p}_m \tag{4.20}$$

Eq. 4.20 can be used as an estimator of system reliability. This equation contains two approximations not contained in the exact equation for the probability of network failure, given by Eq. 4.16. These approximations are 1) the use of $\hat{p}_m$ to estimate $P(F \mid E_M)$ and 2) the use of N_{max} for the maximum size removal set, rather than the total number of links in the network, ML, as discussed in relation to Eq. 4.6. These two approximations lead to error in the estimate of network reliability. In addition, the use of Eq. 4.16 assumes that either $P(E_m)$ is known a priori or is estimated using statistical techniques. It should be noted that most previous literature studies assume that component failure probabilities are known, e.g., Ball (1980), Ball and Nemhauser (1983), Le and Li (1989), and Nels and Colbourne (1990). Such an assumption is a convenient simplification of the problem, but is unrealistic for water distribution networks. Statistical estimation of $P(E_m)$ on the other hand is more complicated, but far more realistic. As a result, the use of Eq. 4.20 to estimate network reliability actually introduces three sources of error, the two related to the use of $\hat{p}_m$ and N_{max} discussed earlier and an additional issue related to estimation of the actual failure rates of the components. The means of addressing these three types of errors are discussed in the following section.

4.3 Error in the Estimate of the Probability of Network Failure

Consider each of the three error terms mentioned above separately. The calculation of $\hat{p}_m$ is performed using Monte Carlo simulation. For each trial, in the simulation, the removal of m links either causes the network to fail or does not cause the network to fail. The trials themselves are performed so that they are independent and identical. As a result $\hat{p}_m$, the estimator of $P(F \mid E_m)$, follows a binomial distribution and the variance of $\hat{p}_m$ is

$$VAR(\hat{p}_m) = \frac{(\hat{p}_m)(1-\hat{p}_m)}{t_m} \tag{4.21}$$

The confidence interval for $\hat{p}_m$ is based on the standard error and the fact that $\hat{p}_m$ has can be approximated by the normal distribution,for large values of t_m. It can be seen that, for a level of confidence β, the confidence interval is given by

$$CI_{1-\beta} = \hat{p}_m \pm (Z_{(\beta/2)}\sqrt{(\hat{p}_m)((1-\hat{p}_m)/t_m}) \tag{4.22}$$

where

- $CI_{1-\beta}$ is the $1-\beta$ confidence interval for $\hat{p}_m$
- $Z_{\beta/2}$ is the the value of the standard normal variate corresponding to the probability $1-\beta/2$

Eq. 4.22 illustrates the dependence of the width of the confidence interval on the number of trials t_m. (Note that it is assumed that sufficient trials are made to justify the use of the normal approximation to the binomial distribution.)

The second source of error is caused by using only N_{max} terms of Eq. 4.16 to estimate reliability. The error resulting from the use of this term is then equal to the sum of the omitted terms.

$$error = \sum_{m=Nmax+1}^{ML} P(F \mid E_m)P(E_m) \tag{4.23}$$

In general, the direct calculation of Eq. 4.23 is not possible because of difficulty in estimating $P(E_m)$ for large values of m and the computational burden in determining $P(F \mid E_m)$ for large m. It is, however, possible to bound the error. Expressing the error in terms of the definitions of removal sets and network failure in terms of set notation gives

$$(F \cap (E_{Nmax+1} \cup E_{Nmax+2} \cup \ldots \cup E_{ML})) \subset$$
$$(E_{Nmax+1} \cup E_{Nmax+2} \cup \ldots \cup E_{ML}) \qquad (4.24)$$

If the set of all possible removal sets is denoted as Ω then

$$(E_0 \cup E_1 \cup E_2 \cup \ldots \cup E_{Nmax}) \cup$$
$$(E_{Nmax+1} \cup E_{Nmax+2} \cup \ldots \cup E_{ML}) = \Omega \qquad (4.25)$$

Note that the events in the first term of Eq. 4.25, E_0 through E_{Nmax}, and the events in the second term of Eq. 4.25, E_{Nmax+1} through E_{ML} are mutually disjoint. If the left hand side of Eq. 4.24 is substituted for the second term in Eq. 4.25 and probabilities of all terms in Eq. 4.25 are taken the following expression results.

$$P((F \cap E_{Nmax+1}) \cup (F \cap E_{Nmax+2}) \cup \ldots \cup (F \cap E_{ML}))$$
$$+P(E_0 \cup E_1 \cup E_2 \cup \ldots \cup E_{Nmax}) \leq 1 \qquad (4.26)$$

Expressing Eq. 4.26 in terms of probabilities and noting the E_m are disjoint events

$$\sum_{m=Nmax+1}^{ML} P(F \cap E_m) \leq 1 - (\sum_{m=0}^{Nmax} P(E_m)) \qquad (4.27)$$

Thus, the error in estimating the probability of failure can be bounded by choosing N_{max} sufficiently large that the right hand side of Eq. 4.27 is less than β.

The third source of error is the statistical error in estimating $P(E_m)$ from historical data. Because the estimate of $P(E_m)$ is used extensively in the equations which bound the probability of failure, it is crucial that this estimate be accurate. In Chapter 6 the errors in the values of $P(E_m)$ are discussed and it is shown that the link reliability is highly dependent on pipe material, pipe size and to a lesser extent, pipe age.

An important point to consider is that all three of these error sources act simultaneously and so must be estimated simultaneously. Bonferroni confidence limits should be used to estimate all quantities. Due to the nature of Bonferroni estimation limits, the accuracy of the reliability estimate can decrease if too many parameters are used to make the estimate.

A limitation of this model is that it assumes that failure probabilities are uniform for all links in the network. In Chapter 5, the model is generalized to allow the use of multiple pipe categories based on pipe material, pipe diameter, and pipe age. The method used to estimate network reliability under these additional considerations is described. In addition, the interaction between network size, network complexity and the accuracy of reliability estimation achievable using this method is discussed.

Chapter 5

A Multiple Component Type Model

The model for estimating network reliability developed in Chapter 4 is based on the assumption that pipes have equal failure probabilities. This assumption is not generally justified for pipes in a water distribution system. In this chapter the model developed in Chapter 4 to estimate network reliability is extended to recognize that pipe failure probabilities are a function of pipe category under an assumption that failure events in different categories are independent. These pipe categories considered in the estimation process are based on pipe diameter, material, and age. (The actual model to estimate pipe failure probabilities is developed in Chapter 6.)

The model presented in Chapter 4 used Eq. 4.20 to estimate network reliability and Eq. 4.27 to bound the error in the network reliability estimate. A number of modifications to the model are necessary to generalize these equations. In particular, the partitioning of the removal space must be modified such that sufficient removal sets are chosen to bound the error in the reliability estimate. It is also necessary to develop a model to estimate the probability of failure for events involving multiple pipe categories. Finally, a method is required to determine the accuracy for the simultaneous estimation of parameters used in the reliability calculation.

The assumption that all links have an equal probability of failure is unrealistic for water distribution systems because of the wide variations found in pipe reliability for different categories of pipe. An alternate method of partitioning the space of removal sets is to recognize the existence of the pipe categories explicitly in the partitioning. Subdividing the removal space, based on both the pipe categories and the number of pipes removed in each pipe category, causes the removal set space to be divided into disjoint subsets that have the property that within each subset the probability of removal is the same for all removal sets. Denote the event that m links from pipe category k are removed as E_{mk}. Ensuring that this subdivision of the removal set space covers the space is complicated by events that are comprised of removal sets involving multiple pipe categories. It should be noted, that the following union of events does not cover the removal set space.

$$\cup_{k=1}^{C} \cup_{m=1}^{ML_k} E_{mk} \tag{5.1}$$

where

- C is the number of pipe categories
- ML_k is the number of pipes in pipe category k

The reason Eq. 5.1 does not ensure complete coverage of all removal sets is that joint removal events that involve multiple pipe categories are possible. Thus, enumeration of the possible removal events requires binations of numbers of links to be removed from the various pipe categories. This full enumeration of all possible combinations requires excessive computational effort. The approach that is used in this thesis is to enumerate only a subset of the possible combinations.

Unlike the equiprobable failure case presented in the previous chapter in this case there is no a priori method of choosing which removal events, i.e. combinations of links are most probable. It is reasonable to assume that, for equiprobable events, the simultaneous occurrence of two events is less likely than the occurrence of only one event. This assumption is not valid for networks such as water supply networks with pipe categories which have widely varying failure probabilities. It is possible that a removal event involving several pipes with high failure probabilities is more likely than an event involving only one very reliable pipe.

The approach used in this thesis to address the problem is based on experience with sampling of the City of Winnipeg pipe break data. (The results of this sampling are presented in detail in Chapter 6.) These sampling results show that the probability of occurrence for a removal set decreases rapidly as the size of the removal set increases. Within a pipe component group, removal sets in which more than one pipe fail were found to be more common than that which would be expected if pipe failures were independent. This increased breakage rate appears to be due to an increased occurrence of pipe failures in the immediate temporal and spatial vicinity of previous failures. In this study, the immediate vicinity of a previous pipe failure is taken to mean within 20 m and 90 days of a previous failure. These boundary values are based on the work of Goulter and Kazemi (1988), (1989) in determining the extent of clustering of breaks in the City of Winnipeg.

Because sampled probabilities of occurrence for removal sets decrease rapidly as removal set size increases, combinations of removal sets used in the model are chosen in the order of increasing numbers of links. The first removal set to be considered is the case in which no links are removed. This event is equivalent to the event that no links are removed from any pipe category. Thus, the probability, that no links are removed, can be expressed as

$$P(E_0) = P(E_{0,1} \cap E_{0,2} \cap \dots E_{0,C})$$
$$= P(E_{0,1})P(E_{0,2}) \dots P(E_{0,C}) \tag{5.2}$$

where

- $E_{0,k}$ is the event that no links are removed, from the kth pipe category.

The values on the right hand side of Eq. 5.2 can be determined from data sampling. It should be noted that the second line of Eq. 5.2 is based on the assumption of independence between failures in different pipe categories mentioned previously.

The next event to be considered is the event that a single failure occurs for a given pipe category k. This event is equivalent to the removal of one pipe from category k and no pipes from any other pipe category. The probability of this event can be expressed as:

$$P(1\ failure\ in\ category\ k) = P(E_{1,k})P(E_{0,1}) \dots P(E_{0,k-1})P(E_{0,k+1}) \dots P(E_{0,C}) \tag{5.3}$$

where

- $E_{0,j}$ is the event that no pipes fail in category j, $j \neq k$
- $E_{1,k}$ is the event that exactly 1 pipe fails in category k

The next type of event to be considered is the failure of multiple links from the same pipe category. This type of event is considered at this point because the dependence of failures from the same pipe category tends to make the occurrence of several failures from a single pipe category much more likely than the occurrence of the same number of failures from different pipe categories. The probability of this type of event can be expressed as:

$$P(m\ failures\ in\ category\ k) =$$
$$P(E_{m,k})P(E_{0,1}) \dots P(E_{0,k-1})P(E_{0,k+1}) \dots P(E_{0,C}) \tag{5.4}$$

where

- $E_{0,k}$ is the event that no pipes fail in category j, $j \neq k$
- $E_{m,k}$ is the event that exactly m pipes fail in category k

The last type of event to be considered is the failure of multiple pipes from several pipe categories. An example of this type of event is the failure of one pipe in category one and two pipes in category two. The probability of this event can be expressed as:

$$\begin{aligned} &P(1\ failure\ in\ category\ 1\ and \\ &\quad 2\ failures\ in\ category\ 2) \\ &= P(E_{1,1})P(E_{2,2})P(E_{0,3})\dots P(E_{0,C}) \end{aligned} \tag{5.5}$$

where

- $E_{0,j}$ is the event that no pipes fail in category j, $j \neq k$
- $E_{m,k}$ is the event that exactly m pipes fail in category k

The number of combinations of this last type of event grows rapidly with the number of failures to be considered and with the total number of pipe categories. As a result it is not possible to include all such combinations of pipe failures in the estimation of network reliability. Instead the objective should be to include a sufficient number of the removal events to cover the failure space to within an acceptable degree of accuracy. A set of removal events which covers the component failure space to within an acceptable degree of accuracy is denoted as a covering set.

Since not all removal events can be considered, the selection of removal events to be included in a covering set requires a certain amount of judgment. The strategy used in this thesis is to choose removal events to be included in the covering set in the following way.

1. The event in which no pipes fail

2. All events in which exactly 1 pipe fails

3. All events in which exactly 2 pipes fail from the same category

All of these removal set events are chosen without regard to the sum of their probabilities of occurrence. At this point, the total probability of occurrence covered is calculated for the above events. The sum of the probabilities of occurrence of all events in these categories is called the coverage. This probability (or coverage) is compared to the required level of accuracy. For example, for a required level of accuracy $\beta = 0.01$, if the coverage is greater than 0.99, then the selection of removal set events would stop. If more removal set events are required to achieve sufficient coverage then the following types of events would be considered.

1. Events in which exactly 1 pipe from each of two categories fail

2. Events in which exactly 3 pipes from the same category fail

3. Events in which exactly 1 pipe from each of three categories fail

4. Events in which exactly 2 pipes from 1 category and one pipe fails from another category fail

5. Events in which exactly 4 pipes from 1 category fails

6. Continue in the same pattern with increasing numbers of failures

The probability of each these events is calculated at each stage to check the coverage for sufficiency.

The removal set events chosen for the covering set following the procedure outlined above are then sequentially numbered in the order that they were chosen. The removal set events in the covering set are denoted $RE_i, i = 1, 2, 3, \ldots, MM$, where MM is the number of removal set events chosen. An example of a removal set event is all component failure events involving the simultaneous failure of two, 200 mm diameter, cast iron pipes which are at least 30 years old .

Eqs. 4.20 and 4.27 can now be generalized to calculate reliability and bound the error in the reliability calculation. The analog to Eq. 4.20 to perform this task is:

$$P(F) = \sum_{i=1}^{MM} P(F \cap RE_i) = \sum_{i=1}^{MM} P(RE_i)P(F \mid RE_i) \tag{5.6}$$

where

- RE_i is the ith removal set event in the covering set
- MM is the number of removal set events included in the covering set

The term $P(F \cap RE_i)$ is estimated using Monte Carlo simulation in a similar fashion to the method proposed in the previous chapter. (The mechanics, of how the simulation is performed are discussed in Chapters 8 and 9.)

The analog to Eq. 4.27 is:

$$\sum_{i=MM+1}^{MT} P(F \cap RE_i) \leq (1 - (\sum_{i=1}^{MM} P(RE_i)) \tag{5.7}$$

where

- MT is the total number of possible removal set events that can occur for the the network

5.1 Simultaneous Estimation

In the worst case, Eqs. 5.6 and 5.7, rely on the simultaneous estimates of the probability of occurrence of removal sets for all possible pipe categories and the conditional probability of occurrence of network failure given the removal of a set of links. The accuracy of these equations depends on the accuracy of the model used to predict component failure probabilities for each pipe category and the accuracy with which it is possible to estimate the conditional failure probability terms $P(F \mid RE_i)$ in Eq. 5.6. In addition, the simultaneous estimation of component failure probabilities and conditional failure probabilities, using Bonferroni confidence limits for multiple pipe categories, decreases the confidence which can be placed in the estimate of any one of the values being estimated.

The approach used to estimate the component failure probabilities is to develop a model to estimate the probability of occurrence for removal sets with 0, 1, 2, and 3 elements for each pipe category. These models depend on four parameters for each pipe category. These parameters are LBI, the average probability of an independent pipe break; $K0$, the probability that a second pipe break occurring within one day and 20 m of an independent pipe break, given that the independent pipe break has occurred; $K1$, the probability that a second pipe break occurs within 20 m and between 1 and 90 days of the original break; and $K2$, a factor assumed to take into account factors which cause a decreased number of days with only one pipe break in a day. These spatial and temporal limits (20 m and 90 days) are the same as those used in Goulter and Kazemi (1988)(1989). Once these parameters have been estimated, the estimation models can be used to estimate the required probabilities for Eqs. 5.6 and 5.7.

A regression approach is used to estimate the parameters for each pipe category. Each of the parameters is estimated as a linear function of pipe length and/or pipe age. Details of the regression approach used to determine the values are given in the following chapter. As in the case for equal probabilities of failure used in the previous chapter Monte Carlo simulation is used to estimate the conditional probability term in Eq. 5.6. Note that each case that is simulated requires the estimation of one conditional probability term, thus increasing the number of parameters to be estimated by one. Since all of the parameters for all pipe categories must be estimated simultaneously, Bonferroni confidence limits are used.

If g parameters are to be estimated simultaneously and the required level of confidence is $1-\alpha$, then the required confidence level for each parameter that is estimated, α_R, is found by solving

$$1 - \alpha_R = 1 - \frac{\alpha}{2g} \tag{5.8}$$

where

- g is the number of parameters to be estimated simultaneously
- α is the required level of significance for the simultaneous estimate of all parameters
- α_R is the required level of significance for the estimation of each parameter

Based on Eq. 5.8, the required level of confidence for each parameter to be estimated is much higher than the required level of confidence for the total estimate. As a result Bonferroni confidence intervals are wider than single variate confidence intervals, for the same level of confidence. The extent to which the number of parameters that must be estimated can be limited is an important determinant of the applicability of this model.

5.2 Confidence Limits for Total Probabilities

Eqs. 5.6 and 5.7 use sums of probabilities to estimate the cumulative probability of network failure and to bound the error in the estimate of the cumulative probability of system failure. The sum of the probabilities of occurrence of removal set events is also used to estimate the total probability of occurrence of the coverage set. The value of these cumulative probabilities can be estimated from the Bonferroni confidence limits calculated for each network. It is possible to estimate the expected value of each cumulative probability by summing the maximum likelihood estimators over all failure cases. Simply summing the lower bounds of the Bonferroni confidence limits to obtain a lower limits or the upper bounds to obtain upper limits for these cumulative probabilities is, however, unduly conservative. Rather than simply sum upper or lower bounds a statistical approach to estimating these probabilities is used. The variance of the total probability of failure can be estimated from the variances of the individual failure cases.

Knowledge of the relationship between the events in the covering set is necessary to determine the variance of the sum of probabilities for the covering set. It should be noted that removal set events which affect a single pipe category include virtually all failure events that can be expected for the category and must effectively add up to one. It is, thus, extremely unlikely that the probabilities of occurrence for all failure events affecting a single pipe category are either over or under estimates. If the estimate of the probability of occurrence of 1 failure is high then the other estimates should be low to compensate. In other words it can be expected that events that affect a single pipe category are negatively correlated. The estimated error from negatively correlated events should be less than the error from independent events. It is also reasonable to assume that events that affect a different pipe categories are independent. Thus the assumption was made that the individual failure cases in the covering set are independent. Under this assumption the variance of the total probability of failure, S^2, is the sum of the variances from each case in the covering set, S_i^2. This expression can be written:

$$S^2 = \sum_{i=1}^{n} S_i^2 \tag{5.9}$$

where

- S^2 is the variance of the cumulative probability of failure
- S_i^2 is the variance of the probability of failure in ith distribution in the covering set
- n = number of cases in the covering set

For each failure case in the covering set the Bonferroni limits give a .99 confidence interval for the expected probability of failure. The standard deviation and, hence, the variance can be estimated from the confidence interval. Note that in general the width of a confidence interval is given by:

$$w = 2k\sigma \tag{5.10}$$

where

- w is the known width of the confidence interval
- σ is the standard deviation
- k is a distribution dependent constant

The parameter k is dependent on the underlying probability distribution. The underlying distributions of the cases in the covering set are in general not known. It is, however, conservative to choose a distribution with as small a k as possible, i.e.

the normal distribution. For a .99 confidence interval for the normal distribution k = 2.576. The standard deviation can then be estimated for each case in the covering set as

$$S_i = \frac{w_i}{2k} = \frac{w_i}{5.15} \tag{5.11}$$

where

- w_i is the width of the Bonferroni confidence interval for the i th distribution in the covering set
- S_i is the standard deviation of the i th distribution in the covering set

The variance of the cumulative probabilities can then be estimated using S^2, as given by Eq. 5.9. The variance S^2 can be used both for the calculation of confidence limits and for hypothesis testing with respect to system reliability. The t-distribution with n-1 degrees of freedom (n is the number of sets in the covering set) and be used for both confidence limits and hypothesis testing.

5.3 Factors Affecting Model Applicability

From Eqs. 5.8 - 5.11 it can be seen that both the width of the Bonferroni Confidence interval w_i and the variance of the cumulative failure probabilities S^2 increase with the number of parameters estimated. Thus, the number of parameters that must be estimated is crucial to the applicability of the model. The number of parameters requiring estimation and thus the applicability of this model is affected by a number of factors. The first factor is the complexity of the system. A system with a large number of different components requires the simultaneous estimation of a large number of parameters.

The second factor that affects the number of parameters to be estimated is related to network size. In general, as the network size increases, the number of simultaneous failures likely to occur increases. The increase in the number of simultaneous failures can cause a large increase in the number of failure combinations that must be considered to achieve coverage of the removal space. It should be noted that for water supply distribution systems there is evidence that the number of simultaneous failures that are likely to occur increases much more slowly than system size. Jacobs and Goulter (1991) found that for the entire City of Winnipeg water distribution network, with almost 2300 km of pipe, more than 20 failures were reported on less than one per cent of all operating days. If the assumption is made that most pipe breaks are repaired on the reporting day this suggests that only a small fraction of all pipe links are inoperative at any one time.

A third factor that affects the number of parameters to be estimated is the mean time to repair. For pipe failures to interact the breaks must be in existence simultaneously. A shorter mean time to repair lessens the probability that breaks can interact and thus lowers the number of simultaneous breaks that must considered.

Finally, the failure rate, of the pipes in the network, must be considered. It is self evident that a higher failure rate leads to more failures (and more simultaneous failures) which, in turn, requires the estimation of more parameters. It should be noted that the failure rate estimates used in this study are based on data from Engineering District 4 in the City of Winnipeg. District 4 has a lower pipe breakage rate than is reported for the City of Winnipeg. Winnipeg with pipe breakage rates that have reached approximately one break per kilometer per year, has one of the highest pipe breakage rates of any major city in either the United States or Canada.

One of the major prerequisites for the development of this model is the availability of accurate data on both the failure history and the pipe inventory for the network being analyzed. The pipe failure data used for this study are contained in a break data base gathered by the Waterworks, Waste and Disposal Department of the City of Winnipeg. These data are reasonably accurate and are therefore appropriate for this study. On the other hand, the computerized information about the pipe inventory was found to be fragmentary and incomplete. Considerable effort to augment and verify the computerized pipe inventory information was made, during the course of the study. Documentation on the pipe inventory is given in Appendix A. The next chapter gives the derivation of the models to estimate pipe component reliability for $m = 0, 1, 2,\ and\ 3$ simultaneous breaks.

Chapter 6

ESTIMATING PROBABILITY OF EXACTLY m BREAKS

The key parameters of the models developed in Chapters 4 and 5 are, $P(E_{mk})$, the probability of occurrence of exactly m simultaneous breaks for the kth pipe category and, N_{max} the maximum number of simultaneous breaks that must be considered to obtain a given level of reliability, N_{max}. No theoretical approaches or a priori data to determine the values of these parameters currently exist.

It is necessary for the pipe failure model to accurately predict failure probabilities for a variety of pipe materials, sizes, ages, and lengths as these variables affect the rate of breakage, represented as the average number of breaks per unit length and therefore have a marked effect on failure behaviour. When water pipe breaks are categorized by pipe material, diameter, and age marked differences in the average breakage rate can be seen (See Table 6.1). The differing average failure rates among these pipe categories imply that a separate function must be used to describe failure behaviour for each pipe category. In this thesis these functions are based on sampling from historical pipe break data for the City of Winnipeg, Canada (This data base is described in Appendix A). As such the values of the parameters are valid strictly only for the City of Winnipeg situation. However, the techniques used to derive these parameters have general applicability to other situations where similar data exists.

6.1 Sampling from the City of Winnipeg Data Base

The actual sampling is performed to determine the number of days in which 0, 1, 2, 3, etc. of each of four classes of failure events occurred for each sample selected. The classes of failure events are:

1. days on which pipe failures occurred
2. days with independent pipe failures
3. days with dependent pipe failures on the same day as a causal independent pipe break
4. days with dependent pipe failures on a day subsequent to causal independent pipe breaks

Detailed definitions of these events are given in Section 6.4.

In the sampling process sections of water mains (links) are selected. These links are segregated into pipe categories defined by the pipe material, pipe diameter, and pipe age. In the sampling process the total length of the links in the samples selected is systematically varied so that parameters can be estimated over a range of pipe lengths.

The link data and break data were originally stored in two separate files: L4F007 (link data) and L4F004 (break data). A detailed description of the verification and

Pipe Category Pipe material, diameter, age (mm) (yr)		Average Breaks per Kilometer per Year
Cast Iron Pipe	150 mm, 0-19 yr	0.42
	150 mm, 20-30 yrs	0.34
	150 mm, over 30 yrs	0.70
	200 mm, 0-30 yrs	0.56
	200 mm, over 30 yrs	0.64
	250 mm, 0-30 yrs	0.52
	250 mm, over 30 yrs	0.58
	300 mm, 0-30 yrs	0.50
	300 mm, over 30 yrs	0.72
Asbestos Cement Pipe	150 mm, 0-30 yrs	0.05
	150 mm, over 30 yrs	0.06
	200 mm, 0-30 yrs	0.20
	200 mm, over 30 yrs	0.70
	250 mm, all ages	0.12
	300 mm, all ages	0.03
Ductile Steel Pipe	150 mm, all ages	0.05
	200 mm, all ages	0.04
	250 mm, all ages	0.10
PVC Pipe	150 mm, 0-30 yrs	0.04

Table 6.1: Average number of breaks per kilometer per year for the City of Winnipeg

correction procedures employed to make these data suitable for sampling is shown in Appendix A. A schematic diagram, of the process used to sample from the combined data base, is shown in Fig. 6.1.

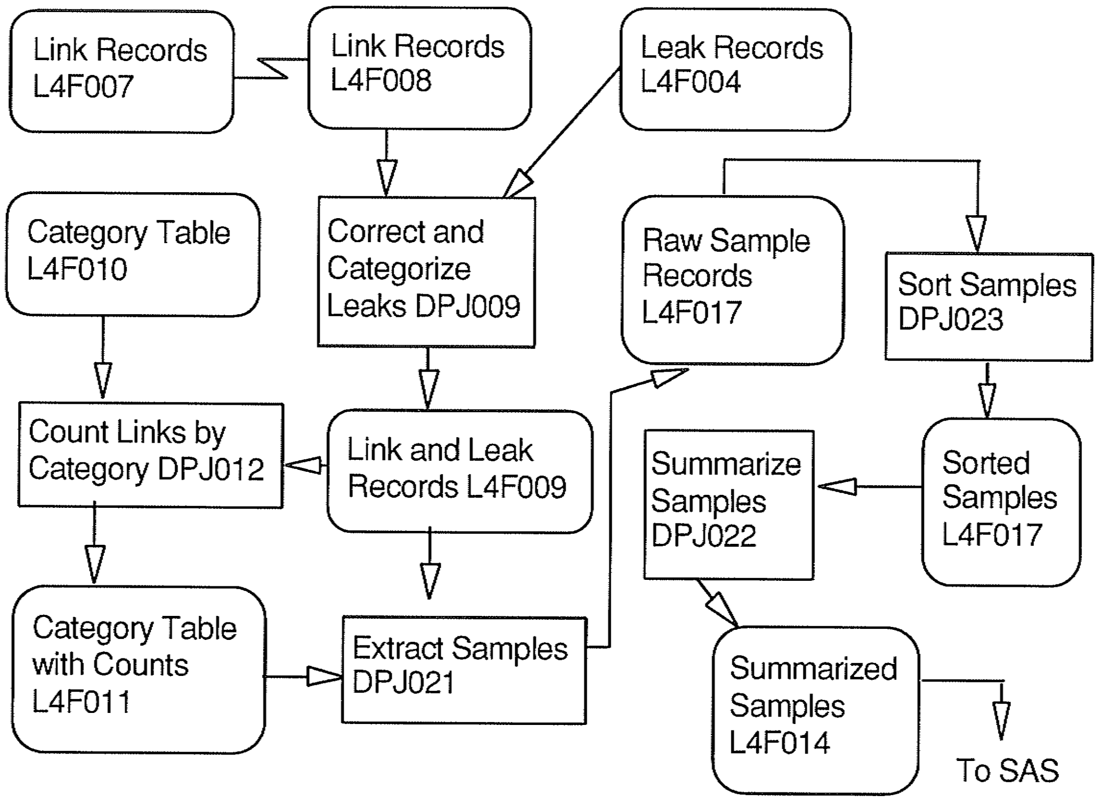


Figure 6.1: Schematic of Computer System to Extract Pipe Breakage Samples

The link data (file L4F007 in Fig. 6.1) were verified and corrected using the process described in Appendix A. The file L4F007 was then transferred to a mainframe computer using a telecommunications link and stored as file L4F008. The program DPJ009 was then run to create a combined database containing both link records and break records. In addition, DPJ009 identifies independent breaks. A break is defined to be independent if it does not occur within 20 m and/or 90 days of a previous break (This definition was used by Goulter and Kazemi (1988)). DPJ009 can create two types of record for each link. The first type of record, the link record, describes the physical characteristics of the link. The link record is always the first record for a link. The second type of record, the break record(s), gives the location and date of occurrence of each break that has been recorded for the link. The next program, DPJ012, counts the number of links in each category where the categories are defined by pipe diameter, pipe material, and pipe age. These counts were used to number, consecutively, the pipe links that form the sampling base for each category. Since the sample period covers the 15 years from 1975 to 1989, pipe age was calculated separately for each year of record.

The third program, DPJ021, randomly selects individual pipe links, from the combined data base. DPJ021 uses a two step approach in sampling data from the combined data base. In the first step, a random number generator is used to select pipe links, without replacement by number, for each category from the sampling base. For each category, 20 samples are taken. The number of links (sample size) in each sample is varied, from an initial value of five links increasing by five links for each successive sample until a maximum sample size of 100 links is reached. The sample sizes were increased in this fashion to vary the total length of pipe in each sample. One sample with each specified number of links is selected. Sampling is performed, sequentially, for each year of record from 1975 to 1989. Since link lengths vary this method does not give precise control of total length of pipe in each sample. The method does, however, provide samples with a range of lengths. It should be noted that because sampled numbers of pipe breaks are compared to pipe length not the number of links in each sample this method does not introduce any inaccuracies in the estimated parameters. For some categories of pipe, there are less links than the required sample sizes. In these cases, all links are sampled.

After the selection of links has been completed for all categories the program reads through the combined data base and creates sample record(s) for each link selected in the previous step. Two types of records can be created for each link. The first record type, the link record, contains data describing the link and contains pipe diameter, pipe material, pipe age, link length, coordinates of the link ends, and sample number. The link record is always the first record created for a link and always contains a code indicating that it is a link record.

The second type of record, the break record(s), describes each break that has occurred on the link. A separate record is created for each break that occurred on the pipe link during the sampling year. These break records contain the coordinates of the break and a code indicating if the break is an independent break. As noted earlier a break is assumed to be independent if it does not occur within 20 m and 90 days of a previous break. The sample records are written to file L4F017.

The next program, DPJ023, sorts the sample records by pipe diameter, pipe material, sample number, pipe age, and break date. The fourth program, DPJ024, summarizes the data for each sample to file L4F014. One record is created for each combination of pipe diameter, pipe material, sample number, and pipe age group. Summary records which contain total pipe length for the sample, number of independent breaks, number of occurrences of days with exactly m = 1, 2, 3, 4, etc. breaks on the same day, independent breaks, dependent breaks which occur on the same day as the causal independent break, and dependent breaks which occur on days subsequent to the causal independent break are then created.

The summary records on file L4F014 are used as the data base for regression analysis to estimate parameters of the distributions used to predict pipe failure rates.

6.2 Variation of Average Breakage Rates

A number of features are of interest with respect to the variation of average breakage rate among the different pipe categories. From Table 6.1 it can be seen that the average number of breaks per kilometer per year varies widely among the various pipe categories. This parameter varies from approximately .7 breaks per kilometer per year to .03 breaks per kilometer, per year. This variance in values of the average number of breaks provides strong evidence that pipe breakage events do not come from a uniform population.

A reasonable supposition in the light of this evidence is that pipe breakage can be expected to increase with age. However, several papers in the literature, e.g., Andreou (1987), O'Day (1982), and Clark and Goodrich (1989), found that pipe failure rates depend only weakly on pipe age and that this weak dependence can be overshadowed by other effects. In particular, for 150 mm cast iron pipe (see Table 6.1) the average number of breaks per unit length does not necessarily increase with age. This result occurs because wall thickness of cast iron pipe has been decreased for newer pipe. The decreased wall thickness has made the newer pipe more prone to failure than the older pipe. However, the increase in breakage rates for newer pipe does not occur for larger pipe sizes. This result is believed to occur because pipe wall thickness in these larger pipes has not been reduced as much as in the smaller diameters because of the greater flows and pressures normally encountered in larger pipes and because of the greater structural strength needed for the pipe itself. These two factors make the larger diameter pipe more resistant to breakage in general.

6.3 Poisson Distribution

Previous work in the literature, e.g., Andreou et al. 1987, Goulter and Coals, 1986, suggests that a Poisson distribution is an adequate model of the rate of pipe failure. For this reason, an attempt was made to fit the Poisson distribution to the City of Winnipeg pipe break data. If the Poisson distribution is used, the average number of breaks per unit length per day is an unbiased estimator of the Poisson parameter lambda (LB). The parameter LB is estimated using the sampled records in L4F014. Regression analysis is performed using the Poisson expression for the probability distribution of days with no breaks. The Poisson expression for the probability distribution of days with no breaks is given in Eq. 6.1.

$$P(E_0) = e^{-LBx} \tag{6.1}$$

where

- $P()$ is the probability function
- E_0 is the event that a day occurs with no independent breaks
- LB is the Poisson parameter
- x is the length of pipe in kilometers

Eq. 6.1 is then transformed to Eq. 6.2 by inverting and taking natural logarithms.

$$\ln \frac{1}{P(E_0)} = LBx \tag{6.2}$$

Recall that the records in file L4F014 contain the observed numbers of days with m=0, 1, 2, 3, etc. breaks. These observations are used to estimate the values of $P(E_0)$ for the samples in L4F014. Eq. 6.2 gives a linear relationship between $\ln \frac{1}{P(E_0)}$

and x. The parameter LB is then estimated as the regression coefficient of this linear relationship using the sampled data in L4F014.

The general equation to estimate the probabilities of days with m= 0, 1, 2, 3, etc. breaks under the assumption that these events follow a Poisson distribution is given by Eq. 6.3.

$$P(E_m) = \frac{(LBx)^m e^{-LBx}}{m!} \tag{6.3}$$

where

- m is the number of breaks that occur in a day
- E_m is the event that a day with m breaks occurs

Note that Eq. 6.1 is a particular case of Eq. 6.3 with m=0.

The value of LB for all age categories of 150 mm cast iron pipe was determined from the data in L4F014 using the previously outlined procedure. The estimated values of LB for the three age categories were then substituted into Eq. 6.3 to estimate the probabilities of occurrence of days with exactly 0, 1, 2 and 3 breaks. Sampling was performed on the City of Winnipeg pipe break data base, for 150 mm cast iron pipe, to estimate the frequency of days with 0, 1, 2 and 3 breaks. Three age categories were used, 0-19 years old, 20-30 years old, and more than 31 years old.

The sampled frequency and the estimated probability of days with 0 breaks for the 0-19 year old 150 mm cast iron pipe category is plotted in Fig. 6.2. Similarly, the sampled frequency, and the estimated probability of occurrence of days with exactly one break, days with exactly two breaks, and days with exactly three breaks are plotted in Figs. 6.3, 6.4, and 6.5, respectively.

Examination of these plots shows that while the Poisson model might provide an adequate fit for exactly 0 simultaneous breaks, it provides a poor fit for both exactly

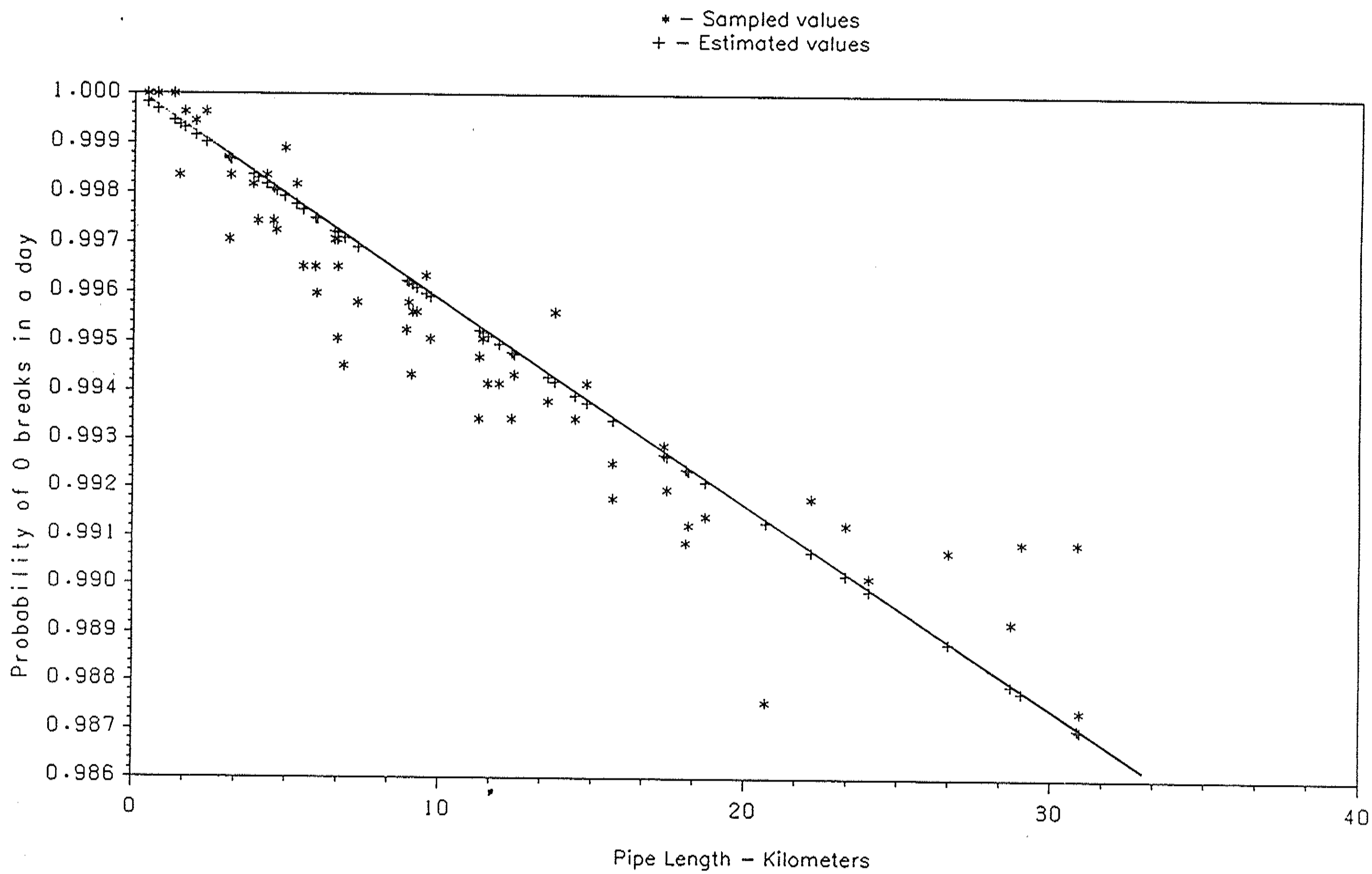


FIGURE 6.2 Poisson estimated probability for m=0 breaks: cast iron pipe, 150mm 0–19 years

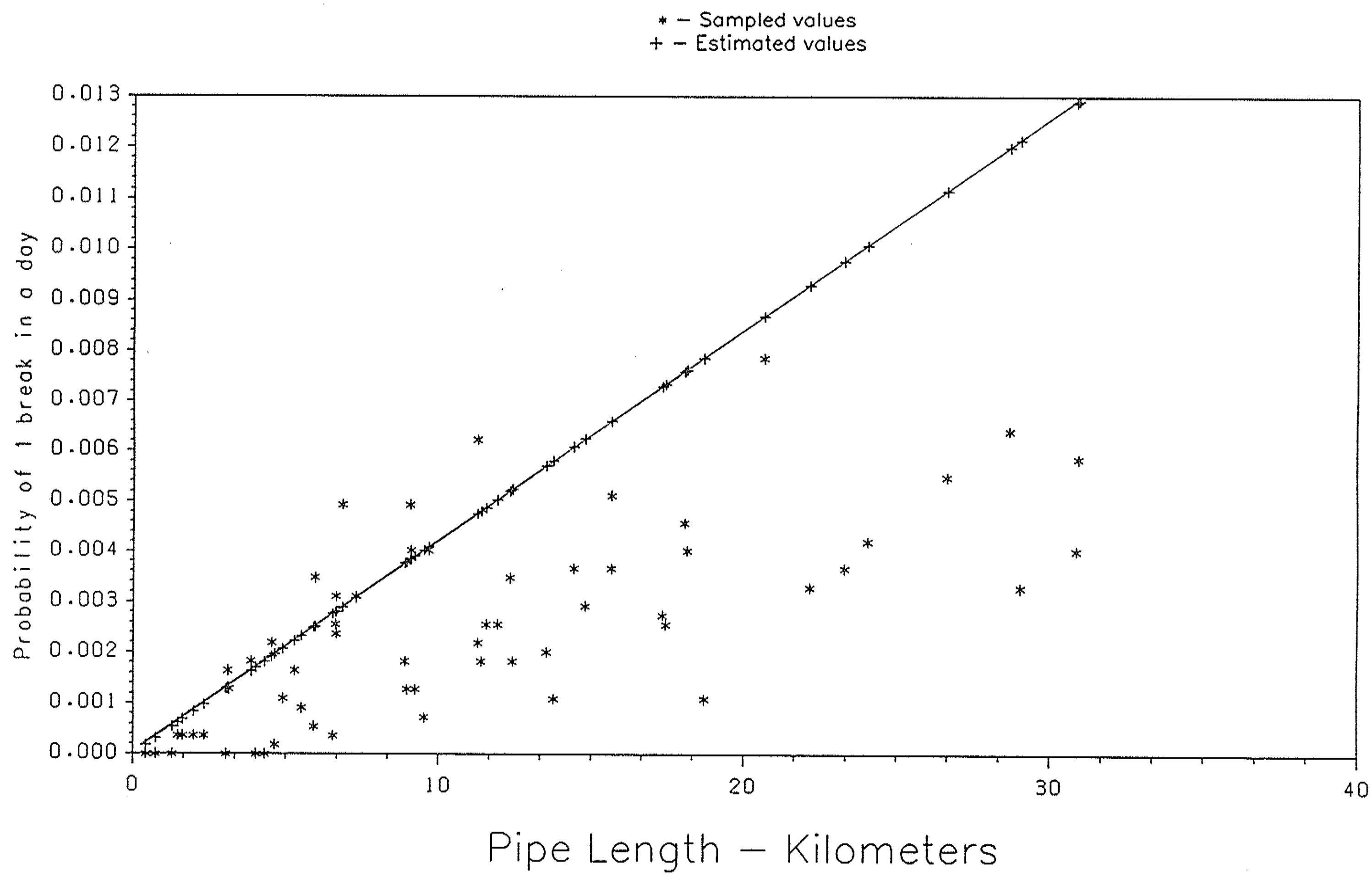


FIGURE 6.3 Poisson estimated probability for m=1 break: cast iron pipe, 150mm 0–19 years

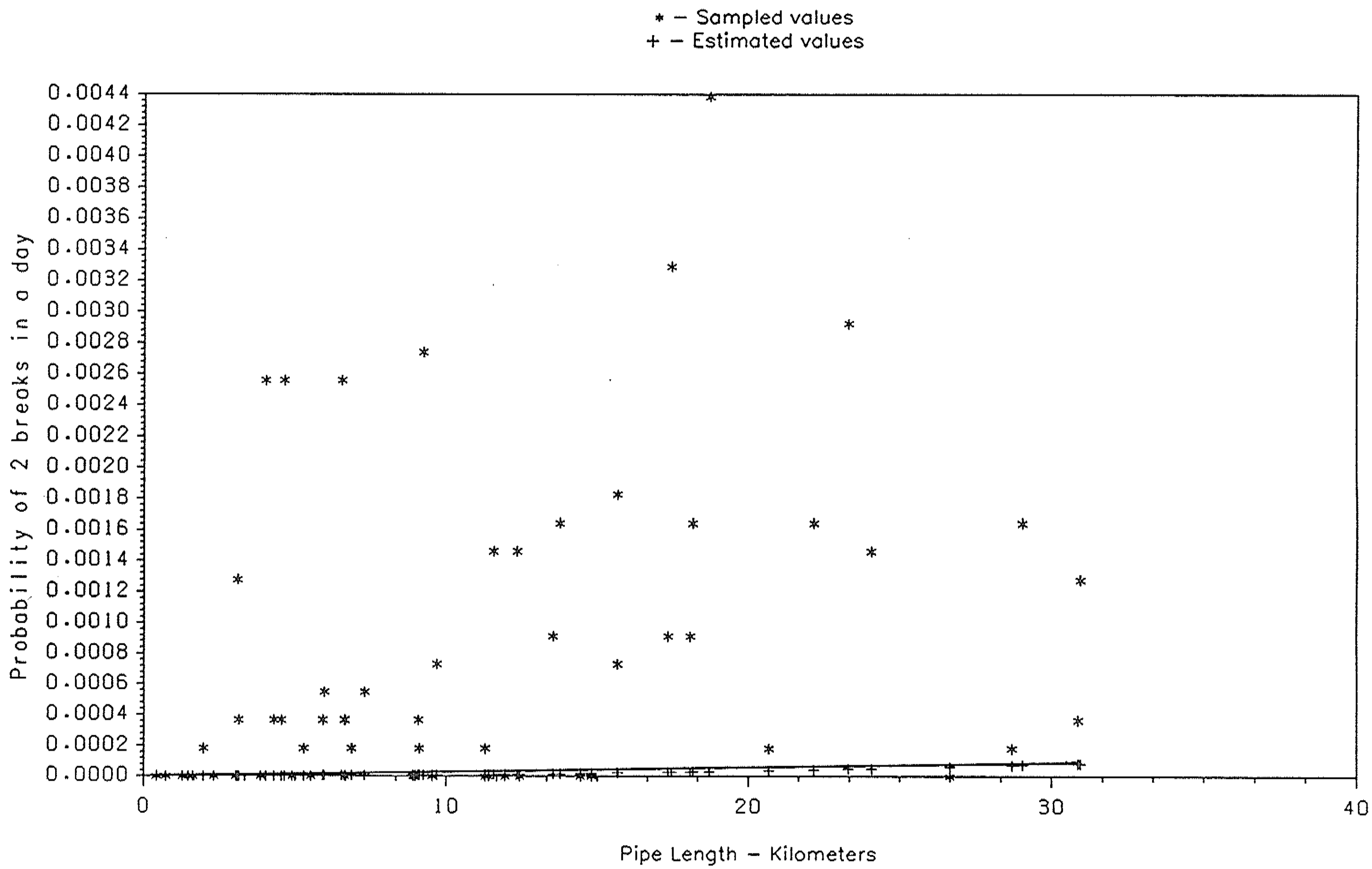


FIGURE 6.4 Poisson estimated probability for m=2 breaks: cast iron pipe, 150mm 0–19 years

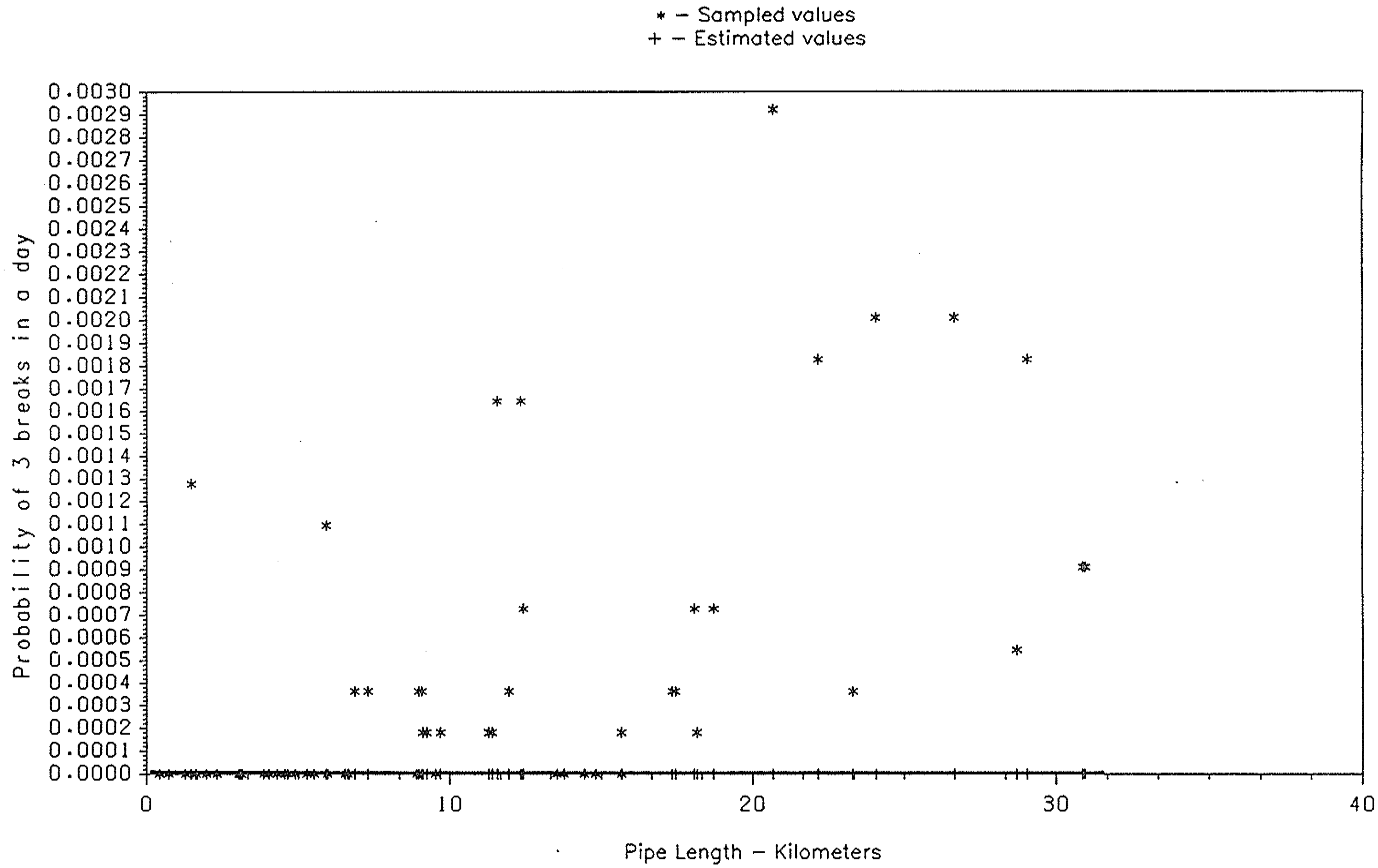


FIGURE 6.5 Poisson estimated probability for m=3 breaks: cast iron pipe, 150mm 0–19 years

one, and exactly two simultaneous breaks. For exactly one break, the Poisson distribution overestimates the probability of occurrence. For exactly two simultaneous breaks, the Poisson distribution underestimates the probability of occurrence. Furthermore, the Poisson distribution predicts almost zero occurrences of three, or more, simultaneous breaks for networks of up to 30 km in length. However, as shown in Fig. 6.5, the actual sampled frequency of days with exactly three breaks is certainly not zero. Similar results occur for the 20-30 year old, and greater than 30 year old categories. (See Appendix B, Figs. B.19 - B.26).

These results provide evidence that the break process in the City of Winnipeg is not a pure Poisson process. However, the relatively good fit achieved for occurrences of zero breaks suggests that the break process is, in some fashion, related to a Poisson process. The assumptions underlying a Poisson process therefore provide a starting point in isolating the deviation of the City of Winnipeg pipe break process from a Poisson process. These assumptions are (following Meyer, 1970):

1. The failure space is divided into a large number of intervals

2. Failure events are independent and have an equal probability of occurrence in any one interval

3. The probability of a failure occurring in any interval is small. The probability of two, or more, failures occurring in any one interval is, as a result, so small that it can be neglected.

In the case of the City of Winnipeg pipe break data, the failure space is divided into 1 day intervals. Breaks that occur on the same day are assumed to be simultaneous. In effect, the failure space for any one year of record is therefore divided into 365 intervals. This level of division is used because it is the level of precision used by the City of Winnipeg to collect the raw data and, to a lesser extent, it gives some recognition of the time required to repair a break. The division of the failure space into 365 intervals is not in direct violation of assumption 1 above (Meyer, 1970).

The work of Goulter and Kazemi (1988) provides insight into the deviation of the performance of the City of Winnipeg pipe network from assumptions 2 and 3. The results of this work show that in Winnipeg there is a significant probability (.146) of more than 1 break occurring within 1 day and 20 m of a previous break. This clustering effect violates assumption 2. In addition, Goulter and Kazemi (1988) show that there is also a significant probability (.064) of a second break occurring within 20 m and 2 to 90 days of an original break. These probabilities are far in excess of what assumption 3 suggests can be expected from a Poisson process. The occurrence of multiple failures in a short time interval (1 day) therefore suggests that the time interval used is too long for the direct use of a Poisson model.

A general lack of independence of failure events has been previously noted in the literature by Le and Li (1989) who note that common failure mechanisms and component capabilities cause failure events to be dependent. In addition, Le and Li (1989) show that, in general, the assumption that failure events are independent is non-conservative. These results suggest that the non-independence of the City of Winnipeg data is typical rather than atypical.

6.4 The Modified Poisson Model

All of the above factors suggest that a model which takes into account the possibility of non-independent breaks is necessary. In developing such a model it is necessary to divide the breaks into three distinct categories, namely,

1. Causal independent break (CIL) - the first break that occurs at a particular location is assumed to be independent of other breaks that occurred before it. It is also assumed that the probability of that break follows a Poisson process.

2. Dependent break occurring on the same day as the CIL (DD) - it is assumed that there is a fixed probability of occurrence for each additional break, that occurs on the same day as the causal independent break. Since DD events are defined to occur within 20 m and on the same day as the CIL it is assumed that the DD events result from localized causes.

3. Dependent break on subsequent day to CIL (DK) - it is assumed that these breaks follow a Poisson distribution, but are not independent of the causal independent breaks. The DK events must occur either at a significant distance (more than 20 m) from a DD event or on a day subsequent to a DD event. It is therefore assumed that these DK events are independent of the DD events.

A modified Poisson model is proposed to predict the probabilities of days with exactly m = 0, 1, 2, and 3 failures. These probability distributions are first estimated as the ratio of the number of occurrences of days with m breaks to the number of days in a year. To derive the modified Poisson distribution five basic events are used to describe the three types of breaks enumerated above. They are:

- I = the event that a day with exactly one CIL occurs
- IA = the event that a day with one or more CIL occurs
- $D0$ = the event that a day with exactly one DD occurs
- $D1$ = the event that a day with exactly one DK occurs
- $D1A$ = the event that a day with one or more DK occurs

In order to get these ratios (distributions) the range of ways in which days with exactly 0, 1, 2, 3 etc. breaks could occur as combinations of the simple events listed above must first be enumerated. Probabilities of these combinations of simple events then are defined symbolically. Finally, mathematical expressions are formulated based on the characteristics of each term in the symbolic probability expression. Thus, four equations are derived for the probability distributions of days with m = 0, 1, 2, 3, etc. breaks. These equations for the probability of occurrence of days with exactly 0, 1, 2, 3, etc. breaks constitute the modified Poisson model. It should be noted that the events in IA and D1A are not assumed to be independent because the underlying CIL and DK events are not independent. Events in D1A are, however, assumed to be independent of events in D0 because because the underlying DD and DK events are assumed to independent.

6.4.1 Probability Distributions for Basic Events

A mathematical expression can be formulated for each of these simple events. These expressions require the estimation of a number of parameters from the data in L4F014. (The actual methods used to estimate these parameters are described in Section 6.5.) As was previously mentioned it is assumed that the independent events I and IA follow a Poisson distribution. Thus LBI is the Poisson parameter for this distribution for a 1 km length of pipe and LBIx is the Poisson parameter for an x km of pipe length. Note that like LB, LBI is the parameter of breaks per day per kilometer.

The general expression for the probability of occurrence of a day with m = 0, 1, 2, 3, etc. independent breaks is thus given by:

$$P(I_m) = \frac{(LBIx)^m e^{-LBIx}}{m!} \tag{6.4}$$

where

- m is the number of independent breaks that occur in a day
- I_m is the event that a day with m independent breaks occurs

In particular, note that the event $\overline{IA}$ is equivalent to the event that a day in which no independent break occurs. The equation for this event is:

$$P(I_0) = e^{-LBIx} \tag{6.5}$$

where I_0 is the event that a day with no independent breaks occurs

A parameter K0 is used to denote the probability of occurrence of D0 events given that a CIL event has occurred. Thus the conditional probability events can be expressed as:

$$P(D0 \mid I) = K0 \tag{6.6}$$

and

$$P(\overline{D0} \mid I) = 1 - K0 \tag{6.7}$$

It should be noted that no event is included which involves multiple DD events with a common CIL event is included in this event as it is assumed that the probability of a day with two or more DD events with a common CIL event is zero. The basis of this assumption is that the occurrence of two or more DD events requires two repairs be performed and three breaks (a CIL, and 2 DD's) to occur within 1 day and in less than 20 m of water main. Such an event, while possible, is extremely improbable.

A parameter K1 is used to estimate D1 and D1A events. Note that D1 and D1A events are defined as days in which DK breaks occur. It is assumed that D1 and D1A events follow a Poisson process with parameter K1x for a pipe length of x km. The general expression for probability of occurrence of days with DK events is thus:

$$P(DK_m) = \frac{(K1x)^m e^{-K1x}}{m!} \tag{6.8}$$

where

- m is the number of DK breaks that occur in a day
- DK_m is the event that a day with m DK breaks occurs

In particular, note that the event $\overline{D1A}$ is equivalent to the event that a day with no DK breaks occurs. The equation for this event is:

$$P(DK_0) = e^{-K1x} \tag{6.9}$$

where DK_0 is the event that a day with no DK breaks occurs

A final parameter K2 is an empirically based adjustment factor. The previously defined parameters LBI, K0, and K1 are defined using the assumption that the probability of occurrence of a pipe break is uniform for all days of the year. Anecdotal evidence (Arnold, 1960) and some general statistical evidence (Goulter and Kazemi, 1988) suggest that the this assumption is not entirely justified. These papers suggest that environmental and operating factors tend to make the occurrence of breaks more likely on some days of the year. The data base on which this study is based does not contain sufficient data to correlate days with pipe breaks with these environmental and operating factors.

The purpose of the parameter K2 is to take into account to some extent the non-uniform distribution of breaks over the days of the year. K2 is calculated as the ratio of observed to predicted days with exactly one independent break. Details of how K2 is used are described in Section 6.4.2 while the method used to calculate K2 is given in Section 6.5.4.

Number of Independent Breaks m	Predicted Probability of Occurrence
1	.025134
2	.0003242
3	.0000027
4	.000000018

Table 6.2: Predicted number of days with m independent breaks for 35 km of 150 mm cast iron pipe older than 30 years

An additional simplifying assumption can be used on the basis of the information shown in Table 6.2. It can be seen in this table that for the lengths of pipe used in this study that the predicted probability of a day with 2 or more independent breaks is quite small. This prediction is based on a "sample" length of 35 km of pipe and the pipe category with the highest pipe breakage rate, 150 mm cast iron pipe older than 30 years. Based on these low failure probabilities all events involving days with two or more independent breaks are assumed to have a zero probability of occurrence.

As was previously mentioned the first step in the derivation of expressions to predict the probability of occurrence of days with m = 0, 1, 2, 3 etc. breaks is the formulation of symbolic expressions for these probability distributions. These symbolic expressions are formulated as combinations of the simple events I, IA, D0, D1, and D1A. The formulas in Eqs. 6.4 - 6.9 and the constant K2 are used to convert these symbolic simple events into mathematical expressions which are use to calculate the probability of occurrence of days with m = 0, 1, 2, and 3 breaks in the following sections.

6.4.2 Probability model for days with m=0 breaks

The equation for the probability of occurrence of a day with exactly 0 breaks is developed first.

The events necessary to define a day with zero breaks occurs are:

1. A day without 1 or more independent breaks (CIL)

2. A day without 1 or more dependent breaks (DK) that occur on days subsequent to the CIL

Note that the event (DD), days with dependent breaks that occur on the same day as an independent break (CIL) has no effect on the number of days with zero breaks. Symbolically, this model is given as:

$$(E_0) = (\overline{IA} \cap \overline{D1A}) \tag{6.10}$$

where E_0 is the event that a day with zero breaks occurs

As mentioned previously, the events IA and D1A are assumed not to be independent. No a priori information exists on the relationship between these events. It should be noted that the underlying breaks for an IA event are CIL's and the underlying breaks for a D1A event are DK's. In terms of system reliability the most conservative assumption would be that DK breaks always occur on a day with CIL breaks because this situation ensures that at least two breaks have occurred on the same day. The assumption that DK breaks only occur on days in which CIL breaks have already ocurred is thus equivalent to the assumption that D1A is a subset of IA. There are some physical factors which support this conservative assumption.

It should be noted that DK breaks are limited by definition to occur within 20 m and 90 days of a previous CIL. Thus, the effective length of pipe which can contribute to a DK break is small compared to the length of pipe which can contribute to a CIL break. Since the CIL and DK breaks underly the IA and D1A events, respectively, it then follows that on any given day a much greater length of pipe contributes to IA events than to D1A events. In addition, it can reasonably assumed that a set of physical causes derived from climatic conditions act to stress the whole network simultaneously. For a D1A event to occur on a day in which no I1A event occurs it would be necessary for these stresses to be great enough to cause a DK break but not great enough to cause a CIL anywhere else in the network. While this scenario is possible it is reasonable to assume that it is unlikely given the much longer length of pipe which on any given day can contribute to IA events. It should be noted that in terms of system reliability this assumption is conservative because it implies that all events in $D1A$ contribute to failure events involving two or more pipes. In spite of being somewhat of an oversimplification this conservative assumption is used to calculated the probability of occurrence of days with no failures.

Using the subset relationship between $D1A$ and IA and converting Eq. 6.10 into probabilities gives:

$$P(E_0) = P(\overline{IA} \cap \overline{D1A}) = P(\overline{IA}) \tag{6.11}$$

Eq. 6.5 can then be used directly to express Eq. 6.11 in terms of LBI. The equation to predict the probability of a day with no breaks becomes:

$$P(E_0) = e^{-LBIx} \tag{6.12}$$

6.4.3 Probability model for days with m=1 break

The probability distribution of days with exactly one break is derived in a similar fashion to the method used to derive the distribution of days with zero breaks. The ways in which one break can occur in a day are:

1. A day with 1 independent break (CIL) and no dependent breaks (DD or DK)

2. A day with 1 dependent break (DK) that occurs on a day subsequent to the CIL and no independent breaks (CIL)

Symbolically these assumptions can be expressed as:

$$E_1 = ((I \cap (\overline{D0} \cap \overline{D1A})) \cup (\overline{IA} \cap D1)) \tag{6.13}$$

Converting Eq. 6.13 into probabilities gives:

$$P(E_1) = P((I \cap (\overline{D0} \cap \overline{D1A})) \cup (\overline{IA} \cap D1)) \tag{6.14}$$

The assumption used in the previous section that the events in $D1A$ form a subset of the events in IA implies that the probability of occurrence of the second term in Eq. 6.14 is zero. This assumption does not reduce the number of DK breaks but implies that the DK breaks occur on days with CIL. Since these DK breaks occur on days with CIL they contribute directly to the number of days with two or more breaks. It can be expected that this contribution is in addition to the number of days with multiple breaks that would otherwise be predicted. No a priori method exists to apportion these extra days with multiple breaks to days with 2, 3, etc. breaks. Several trials were conducted and it was found that apportioning 40% of these days to days with 2 breaks, 30% to days with 3 breaks and 30% to days with more than 3 breaks provided a reasonable fit.

In addition, the use of Eq. 6.12 to estimate the probability of a day with no breaks only provides information on the total number of days in which no breaks occur and provides no information with respect to the probability distributions of days with exactly 1, 2, 3, etc. breaks. It was, however, assumed previously that a physical cause acting over the whole network causes the occurrence of greater numbers of CIL events on some days of the year. The assumption of the clustering of independent failure events on some days of the year is conservative from the point of view of reliability because it implies that more independent events contribute to the occurrence of days with more than one failure.

To take into account this temporal clustering effect for CIL events a constant $K2$ which relates the observed number of days with exactly one CIL event to the predicted number is assumed. The reduction in the number of days with exactly 1 CIL by a factor K2 cannot be assumed to reduce the total number of days with breaks. These CIL are instead redistributed to days with multiple breaks and so contribute directly to the number of days with two or more breaks. It can be expected that this contribution is in addition to the number of days that would otherwise be predicted. As with the DK breaks no a priori method exists to apportion these extra days with multiple breaks to days with 2, 3, etc. breaks. Several trials were conducted and it was again found that apportioning 40% of these days to days with 2 breaks, 30% to days with 3 breaks and 30% to days with more than 3 breaks provided a reasonable fit.

This redistribution of the probability of days with one break to days with multiple breaks is admittedly an inexact method of taking into account the tendency of breaks to be non-uniformly distributed over the days of the year. The recognition of this non-uniform distribution has, however, been recognized in the literature only in an anecdotal fashion (see Arnold, 1960) or using very broad statistical measures such as the variation of the number of breaks per month over the months of the year (Goulter and Kazemi, 1988).

Noting that the second term of Eq. 6.14 is zero and expressing $\overline{D0}$ in terms of conditional probabilities Eq. 6.14 becomes:

$$P(E_1) = P(I \cap \overline{D1A})P(\overline{D0} \mid (I \cap \overline{D1A})) \tag{6.15}$$

Recall that D0 is assumed to independent of D1A so Eq. 6.15 becomes:

$$P(E_1) = P(I \cap \overline{D1A})P(\overline{D0} \mid I) \tag{6.16}$$

Eqs. 6.4 - 6.9 and the constant K2 can now be used to express Eq. 6.16 in mathematical terms:

$$P(E_1) = LBIxe^{-LBIx}e^{-K1x}(1 - K0)K2 \tag{6.17}$$

6.4.4 Probability model for days with m=2 breaks

The development of the probability distribution for the days with 2 breaks is similar to the method used for days with m = 0 and 1 breaks. The ways considered in this thesis in which a day with two breaks can occur are:

1. A day with 2 independent breaks (CIL) and no dependent breaks (DD or DK)
2. A day with 1 independent break (CIL) and no dependent breaks (DD) and 1 dependent break (DK)
3. A day with 2 dependent breaks (DK) and no independent breaks (CIL)
4. A day with 1 independent break (CIL), 1 dependent break (DD), and no dependent breaks (DK)
5. A day with 2 dependent breaks (DD), no dependent breaks (DK) and no independent breaks (CIL)

It should be noted that item 5 requires the occurrence of a day with two DD events. It is assumed that the probability of occurrence of a day with two or more DD events is zero. By definition of a DD event a CIL must occur on the same day and within 20 m of the DD event. The occurrence of 2 DD's requires at least one CIL to occur. Therefore, the occurrence of two or more DD events requires two repairs be performed and three breaks (a CIL, and 2 DD's) occur within 24 hours and in less than 20 m of water main. Such an event, while possible, is extremely improbable. No term for item 5 is included in the following equation which expresses these assumptions symbolically.

$$\begin{aligned}(E_2) = (I \cap (\overline{D0} \cap \overline{D1A}) \cap (I \cap (\overline{D0} \cap \overline{D1A})) \cup \\ (I \cap (\overline{D0} \cap D1)) \cup \\ (\overline{IA} \cap D1) \cap (\overline{IA} \cap D1) \cup \\ (I \cap (D0 \cap \overline{D1A}))\end{aligned} \tag{6.18}$$

Converting Eq. 6.18 into probabilities gives:

$$\begin{aligned}P(E_2) = P(I \cap (\overline{D0} \cap \overline{D1A}) \cap \\ I \cap (\overline{D0} \cap \overline{D1A})) + \\ P(I \cap (\overline{D0} \cap D1)) + \\ P(\overline{IA} \cap D1) \cap (\overline{IA} \cap D1) + \\ P(I \cap (D0 \cap \overline{D1A}))\end{aligned} \tag{6.19}$$

The first term of Eq. 6.19 can be simplified by noting that

$$(I \cap (\overline{D0} \cap \overline{D1A}) \subset I \tag{6.20}$$

As a result of Eq. 6.20 the following relationship holds:

$$(I \cap (\overline{D0} \cap \overline{D1A}) \cap (I \cap (\overline{D0} \cap \overline{D1A}) \subset (I \cap I) \tag{6.21}$$

Converting Eq. 6.21 to probabilities gives:

$$P(I \cap (\overline{D0} \cap \overline{D1A}) \cap (I \cap (\overline{D0} \cap \overline{D1A})) \leq P(I \cap I) \tag{6.22}$$

Based on the values in Table 6.2 it can be seen that the right hand side of Eq. 6.22 is approximately zero and, therefore, the first term of Eq. 6.19 can be neglected.

It should be further noted that the subset relationship between D1A and I implies that:

$$\overline{IA} \cap D1 = \emptyset \tag{6.23}$$

Since the probability of occurrence of $\emptyset$ is zero this relationship implies that the third term of Eq. 6.19 is zero. Applying these simplifications to Eq. 6.19 represented by these interpretations of Eq. 6.22 and 6.23 gives:

$$\begin{aligned} P(E_2) = P(I \cap (\overline{D0} \cap D1)) + \\ P(I \cap (D0 \cap \overline{D1A})) \end{aligned} \tag{6.24}$$

Expressing D0 and $\overline{D0}$ in terms of conditional probabilities gives Eq. 6.25:

$$\begin{aligned} P(E_2) = P(I \cap D1)P(\overline{D0} \mid (I \cap D1)) + \\ P(I \cap \overline{D1A})P(D0 \mid (I \cap D1A)) \end{aligned} \tag{6.25}$$

Recall that D0 is assumed to be independent of D1A and so Eq. 6.25 becomes:

$$P(E_2) = P(I \cap D1)P(\overline{D0} \mid I) + P(I \cap \overline{D1A})P(D0 \mid I) \tag{6.26}$$

Eq. 6.26 can now be converted to a mathematical form on a term by term basis using Eqs. 6.4- 6.9 the constant K2 giving:

$$\begin{aligned} P(E_2) = ((LBIx)e^{-LBIx})(K1x)e^{-K1x}(1 - K0) + \\ ((LBIx)e^{-LBIx})(e^{-K1x})K0 + \\ 0.4(LBIxe^{-LBIx})(1 - K0)(1 - K2)(e^{-K1x}) + \\ 0.4(e^{-LBIx})(K1x)e^{-K1x} \end{aligned} \tag{6.27}$$

Note that the last two terms of Eq. 6.27 are not listed in Eq. 6.26. These terms come from the probabilities transferred from days with 1 break discussed previously.

6.4.5 Probability model for days with m=3 breaks

The derivation of the probability distibution for days with m=3 breaks follows the method used for days with two breaks. The ways considered in this thesis in which three breaks can occur in one day are:

1. A day with 3 independent breaks (CIL) and no dependent breaks (DD or DK)
2. A day with 2 independent breaks (CIL) and 1 dependent break (DK) and no dependent breaks (DD)
3. A day with 2 independent breaks (CIL) and 1 dependent break (DD) and no dependent breaks (DK)
4. A day with 1 independent break (CIL) and 2 dependent breaks (DK)
5. A day with 1 independent break (CIL), 1 dependent break (DD), and 1 dependent break (DK)

Note that cases involving more than 1 dependent break on the same day (D0) or cases involving $(\overline{IA} \cap D1)$ have not been included because of the assumption that the probability of occurrence of these events is zero.

Symbolically this combination of the way in which three breaks can occur is given as:

$$\begin{aligned}
E_3 = (I \cap (\overline{D0} \cap \overline{D1A})) \cap (I \cap (\overline{D0} \cap \overline{D1A})) \cap \\
(I \cap (\overline{D0} \cap \overline{D1A})) \cup \\
(I \cap \overline{D0}) \cap (I \cap \overline{D0}) \cap (\overline{IA} \cap D1) \cup \\
(I \cap \overline{D0} \cap D1 \cap D1) \cup \\
(I \cap D0 \cap D1)
\end{aligned} \tag{6.28}$$

Converting Eq. 6.28 into probabilities gives:

$$P(E_3) = P((I \cap (\overline{D0} \cap \overline{D1A})) \cap (I \cap (\overline{D0} \cap \overline{D1A})) \cap$$
$$(I \cap (\overline{D0} \cap \overline{D1A}))) +$$
$$P((I \cap \overline{D0}) \cap (I \cap \overline{D0} \cap D1)) +$$
$$P(I \cap \overline{D0} \cap D1 \cap D1) +$$
$$P(I \cap D0 \cap D1) \tag{6.29}$$

Based on the relationship given in Eq. 6.20 the following relationship holds:

$$(I \cap (\overline{D0} \cap \overline{D1A})) \cap (I \cap (\overline{D0} \cap \overline{D1A})) \cap$$
$$(I \cap (\overline{D0} \cap \overline{D1A})) \subset (I \cap I \cap I) \tag{6.30}$$

Converting Eq. 6.30 to probabilities gives:

$$P((I \cap (\overline{D0} \cap \overline{D1A})) \cap (I \cap (\overline{D0} \cap \overline{D1A})) \cap$$
$$(I \cap (\overline{D0} \cap \overline{D1A}))) \leq P(I \cap I \cap I) \tag{6.31}$$

It can thus be seen that based on Eq. 6.31 and the values in Table 6.2 the first term of Eq. 6.29 can again be neglected. Next consider the second term of Eq. 6.29. Again based Eq. 6.20 the following relationship holds:

$$(I \cap \overline{D0}) \cap (I \cap \overline{D0}) \cap D1) \subset (I \cap I) \tag{6.32}$$

Converting Eq. 6.32 into probabilities gives:

$$P((I \cap \overline{D0}) \cap (I \cap \overline{D0}) \cap D1) \subset P(I \cap I) \tag{6.33}$$

Thus, based on the values in Table 6.2 and Eq. 6.33 the second term of Eq. 6.29 can also be neglected giving:

$$P(E_3) = P((I \cap \overline{D0}) \cap D1 \cap D1) + P(I \cap D0 \cap D1) \tag{6.34}$$

Expressing D0 in terms of conditional probabilities Eq. 6.34 becomes:

$$\begin{aligned} P(E_3) = {} & P(I \cap D1 \cap D1)P(\overline{D0} \mid (I \cap D1 \cap D1)) + \\ & P(I \cap D1)P(D0 \mid (I \cap D1)) \end{aligned} \tag{6.35}$$

Using the independence of D1 and D0 Eq. 6.35 becomes:

$$P(E_3) = P(I \cap D1 \cap D1)P(\overline{D0} \mid I) + P(I \cap D1)P(D0 \mid I) \tag{6.36}$$

Expressing Eq. 6.36 using Eqs. 6.4-6.9 and the constant K2 gives:

$$\begin{aligned} P(E_3) = {} & ((LBIx)e^{-LBIx}(K1x)e^{-K1x})^2(1 - K0) + \\ & (LBIxe^{-LBIx})K1(K0) + \\ & 0.3(LBIxe^{-LBIx})(1 - K0)(1 - K2)(e^{-K1x}) + 0.3(e^{-LBIx})(K1x)e^{-K1x} \end{aligned} \tag{6.37}$$

Note that the last two terms of Eq. 6.37 are not listed in Eq. 6.36. These terms come from the probabilities transferred from days with 1 break discussed previously.

Similar expressions could be derived for the modified Poisson distribution for greater values of m, in a manner similar to that used for m=2 and m=3 breaks above. However, it was found that three breaks at a time provide sufficient accuracy for the purposes of this study.

6.5 Fitting the Modified Poisson Distribution

The use of the Modified Poisson distribution described in Eqs. 6.12, 6.17, 6.27, and 6.37 requires the estimation of the four parameters, LBI, K0, K1, and K2 in order.

The methods of estimating the values of these four parameters are described in this section.

6.5.1 Estimation of LBI

LBI is estimated from the sampled number of days with independent breaks. Recall that an independent break was defined as a break located more than 20 m and/or occurring more than 90 days after any previous break. The number of days in which 0, 1, 2, 3, etc. independent breaks was sampled directly from the break data base. The assumption that independent breaks are distributed as a Poisson distribution is then used to estimate LBI as a regression coefficient. Note that for a Poisson distribution

$$P(I_0) = e^{-LBIx} \tag{6.38}$$

where

- $P()$ is the probability function
- I_0 is the event that a day occurs with no independent breaks
- x is the length of pipe in kilometers

Inverting both sides of Eq. 6.38, and taking natural logarithms gives:

$$\ln \frac{1}{P(I_0)} = LBIx \tag{6.39}$$

The quantity $P(I_0)$ is estimated from the sampled data as the ratio of the number of days with no independent breaks to the number of days in a year. The coefficient LBI in Eq. 6.39 can be estimated by performing linear regression of the left hand side of Eq. 6.39 on pipe length. Recognizing the fact that breaks are impossible in a zero length of pipe (no pipe) the regression line was forced to pass through the origin.

This regression analysis provides a sufficiently good fit for the estimation of LBI for all pipe categories. For each of the cases examined the coefficients used to estimate LBI, the associated coefficient of variation, and the standard deviation of LBI for each pipe category are shown in Table 6.3. These values are derived from the regression analysis.

Neter et al. (1985) give the standard deviation of the regression coefficient $b1$ for univariate linear regression in which the regression line must pass through the origin as:

$$S(b_1) = \sqrt{\frac{\sum(Y_i - b_1 X_i)^2}{(n-1)\sum(X_i)^2}} \tag{6.40}$$

where

- $S(b_1)$ = standard deviation of the regression parameter $b1$.
- Y_i = ith observed value of the dependent variable.
- b_1 = estimated regression coefficient.
- X_i = ith observed value of the independent variable.
- n = number of observations.

Eq. 6.40 is used to calculate the standard deviation of the parameters estimated using regression analysis on the data in file L4F014. The file L4F014 contains pipe characteristics and break history for each pipe segment in the study area.

6.5.2 Estimation of Parameter K0

K0 is estimated as the ratio between the expected number of days that dependent breaks (DD) occur on the same day as the CIL and the expected number of days

Pipe Category Pipe material, diameter, age (mm) (yr)		Regression Parameters LBI^a	r^2	Standard Deviation of LBI^a
Cast Iron Pipe	150 mm, 0-19 yr	360	.9542	16.5
	150 mm, 20-30 yrs	549	.9609	19.04
	150 mm, over 30 yrs	737	.8483	41.49
	200 mm, 0-30 yrs	516	.8401	2.51
	200 mm, over 30 yrs	597	.8541	3.28
	250 mm, 0-30 yrs	478	.9022	21.7
	250 mm, over 30 yrs	603	.9138	24.8
	300 mm, 0-30 yrs	419	.8205	25.4
	300 mm, over 30 yrs	352	.7549	30.1
Asbestos Cement Pipe	150 mm, 0-30 yrs	79.6	.6386	5.43
	150 mm, over 30 yrs	104	.6608	8.61
	200 mm, 0-30 yrs	181	.3478	25.29
	200 mm, over 30 yrs	699	.9742	29.69
	250 mm, all ages	46.5	.5447	16.28
	300 mm, all ages	58.2	.7352	6.13
Ductile Steel Pipe	150 mm, all ages	91.8	.8658	5.41
	200 mm, all ages	40.8	.5432	6.21
	250 mm, all ages	91.8	.7822	6.11
PVC Pipe	150 mm, 0-30 yrs	26.8	.9542	11.3

a. Values shown increased by a factor of 10^6

Table 6.3: Estimation of parameter LBI

with independent breaks (CIL). It is assumed that the days with dependent breaks (DD) are distributed using a Poisson distribution with parameter KDD, but are not independent of the days with CIL. The actual number of days with dependent breaks(DD) is determined from data sampled from the combined break data base. The expected number of days with dependent breaks is estimated using an equation with the same form as Eq. 6.38.

The Poisson expression for the probability of occurrence of a day with no DD breaks for a pipe length of x km is:

$$P(DD_0) = e^{-KDDx} \tag{6.41}$$

where

- $P()$ is the probability function
- DD_0 is the event that a day occurs with no dependent breaks on the same day as the CIL
- KDD is the Poisson parameter for days with these dependent breaks
- x is the length of pipe in kilometers

Then inverting both sides of Eq. 6.41 and taking natural logarithms gives:

$$\ln \frac{1}{P(DD_0)} = KDDx \tag{6.42}$$

The quantity $P(DD_0)$ is estimated from the sampled data as the ratio of the number of days with no dependent breaks (DD) to the number of days in a year. The coefficient KDD in Eq. 6.42 can be estimated by performing linear regression of the left hand side of Eq. 6.42 on pipe length. The regression line was again forced to pass through the origin because breaks cannot occur on a zero length of pipe.

For each of the cases examined, the coefficients used to estimate KDD, the associated coefficient of variation, and the standard error for each pipe category derived from the regression analysis are shown in Table 6.4. It can be seen from the results in this table that the regression analysis provides a reasonably good fit for the estimation of KDD for most pipe categories. It should be noted that r^2 values were found to be statistically significant only for asbestos pipe in the 200 mm over 30 years old and 250 mm all age categories. It is possible that the failure process in asbestos cement pipe is quite different from the process which occurs in cast iron or ductile steel pipe. No inferences can be made about the value of $K0$ for these pipe categories and $K0$ is shown with a nominal value of zero for these pipe categories. Caution should be used in applying this model to estimate component failure probabilities for asbestos cement pipe.

It is also interesting to note that while ductile steel, and particularly PVC pipe categories have quite low probabilities of days with independent breaks the probability of a second break occurring on the same day and in close proximity is quite high. The parameter K0 can then be estimated using:

$$K0 = \frac{KDDx}{LBIx} = \frac{KDD}{LBI} \tag{6.43}$$

A summary of the values of KDD, LBI, and K0 obtained for each pipe category are shown in Table 6.5.

6.5.3 Estimation of Parameter K1

K1 is estimated from the sampled number of days with dependent breaks (DK) that occur on days subsequent to the CIL. The number of days with 0, 1, 2, 3, etc. dependent breaks (DK) was sampled directly from the break data base. The assumption that these breaks are distributed as a Poisson distribution is then used to estimate

Pipe Category Pipe material, diameter, age	(mm) (yr)	Regression Parameters KDD^a	r^2	Standard Deviation of KDD^a
Cast Iron Pipe	150 mm, 0-19 yrs	54.3	.7493	6.87
	150 mm, 20-30 yrs	88.4	.8977	5.18
	150 mm, over 30 yrs	57.5	.7000	4.90
	200 mm, 0-30 yrs	69.2	.7936	1.17
	200 mm, over 30 yrs	56.8	.7968	1.29
	250 mm, 0-30 yrs	73.6	.8468	2.16
	250 mm, over 30 yrs	67.2	.7288	2.31
	300 mm, 0-30 yrs	24.8	.6944	.818
	300 mm, over 30 yrs	12.4	.4676	.341
Asbestos Cement Pipe	150 mm, 0-30 yrs	0	.2071	2.85
	150 mm, over 30 yrs	0	.0000	
	200 mm, 0-30 yrs	0	.1114	1.53
	200 mm, over 30 yrs	22.8	.7643	3.35
	250 mm, all ages	7.8	.3384	4.31
	300 mm, all ages	0	.0000	
Ductile Steel Pipe	150 mm, all ages	22.0	.6891	2.20
	200 mm, all ages	29.0	.6648	3.49
	250 mm, all ages	31.6	.8021	2.13
PVC Pipe	150 mm, 0-30 yrs	22.5	.5505	11.34

a. Values shown increased by a factor of 10^6

Table 6.4: Estimation of parameter KDD

Pipe Category Pipe material, diameter, age (mm) (yr)		KDD^a	LBI^a	K0
Cast Iron Pipe	150 mm, 0-19 yrs	54.3	360	.151
	150 mm, 20-30 yrs	88.4	549	.161
	150 mm, over 30 yrs	57.5	737	.078
	200 mm, 0-30 yrs	69.2	516	.134
	200 mm, over 30 yrs	56.8	597	.095
	250 mm, 0-30 yrs	73.6	478	.154
	250 mm, over 30 yrs	67.2	603	.111
	300 mm, 0-30 yrs	24.8	419	.059
	300 mm, over 30 yrs	12.4	352	.035
Asbestos Cement Pipe	150 mm, 0-30 yrs	0	79.6	0
	150 mm, over 30 yrs	0	104	0
	200 mm, 0-30 yrs	0	181	0
	200 mm, over 30 yrs	22.8	699	.033
	250 mm, all ages	7.8	46.5	.168
	300 mm, all ages	0	58.2	0
Ductile Steel Pipe	150 mm, all ages	22.0	91.8	.240
	200 mm, all ages	29.0	40.8	.711
	250 mm, all ages	31.6	91.8	.344
PVC Pipe	150 mm, 0-30 yrs	22.5	26.8	.839

a. Values shown increased by a factor of 10^6

Table 6.5: Summary of value of KDD, LBI, and K0

K1 as a regression coefficient. The actual estimation of the value of K1 is performed using an equation of the same form as Eqs. 6.38 and 6.41, namely,

$$P(DK_0) = e^{-K1x} \tag{6.44}$$

where

- $P()$ is the probability function
- DK_0 is the event that a day occurs with no dependent breaks on a day subsequent to the CIL
- x is the length of pipe in kilometers

Inverting both sides of Eq. 6.44 and taking natural logarithms gives:

$$\ln \frac{1}{P(DK_0)} = K1x \tag{6.45}$$

For each of the cases examined the coefficients used to estimate K1, the associated coefficient of variation, and the standard error for each pipe category are shown in Table 6.6. It can be seen from the results in this table that the regression analysis provides a sufficiently good fit for the estimation of K1 for most pipe categories. It should be noted that r^2 values were found to be statistically significant only for asbestos pipe in the 200 mm, over 30 years old category and for the ductile steel, 150 mm all ages and 250 mm diameter all ages pipe categories.

It is possible that the failure process in asbestos cement pipe is quite different from the process which occurs in cast iron or ductile steel pipe. No inferences can be made about the value of $K1$ for these pipe categories and $K1$ is shown with a nominal value of zero for these pipe categories. Caution should be used in applying this model to estimate component failure probabilities for asbestos cement pipe.

No explanation was found for the low r^2 value found for the 200 mm all ages ductile steel pipe category. It is speculated that this pipe was installed with a concrete coating and, thus, might be expected to behave differently from other ductile steel and cast iron pipe (Pers. Comm. Glen Chambers, City of Winnipeg, 1992). Again a nominal value of $K1 = 0$ is assumed.

6.5.4 Estimation of Parameter K2

K2 is estimated from the observed probability of occurrence of days with 1 CIL and the Poisson distribution for days with exactly one independent break. The Poisson parameter for days with exactly one CIL is taken to be LBI. Consistent with the concepts used to derive the values of K0 the expression for the distribution of days with exactly one independent break as given by Eq. 6.4, with m = 1, is given by:

$$P(I_1) = (LBIx)e^{LBIx} \tag{6.46}$$

where

- $P(I_1)$ is the sampled ratio of days with exactly 1 CIL to the number of days in a year
- LBI is the Poisson parameter estimated in Eq. 6.39
- x = the sampled length of pipe

If the probability of occurrence of pipe breaks is equal for all days of the year Eq. 6.46 should be a good predictor of the probability of occurrence of days with exactly one CIL. An indication of the extent to which the probability of occurrence of pipe breaks is unequally distributed over the days of the year, i. e., the extent to which

Pipe Category Pipe material, diameter, age (mm) (yr)		Regression Parameters $K1^a$	r^2	Standard Deviation of $K1^a$
Cast Iron Pipe	150 mm, 0-19 yrs	50.4	.8535	4.53
	150 mm, 20-30 yrs	51.6	.9001	2.98
	150 mm, over 30 yrs	71.7	.7112	6.18
	200 mm, 0-30 yrs	101.0	.6764	7.80
	200 mm, over 30 yrs	72.9	.7595	5.47
	250 mm, 0-30 yrs	97.4	.8719	4.98
	250 mm, over 30 yrs	67.2	.8539	4.13
	300 mm, 0-30 yrs	40.8	.6991	3.63
	300 mm, over 30 yrs	22.5	.4997	3.30
Asbestos Cement Pipe	150 mm, 0-30 yrs	0	.2330	.95
	150 mm, over 30 yrs	0	.0819	.86
	200 mm, 0-30 yrs	0	.1724	3.83
	200 mm, over 30 yrs	93.1	.9133	7.66
	250 mm, all ages	0	.0000	0
	300 mm, all ages	0	.2207	1.06
Ductile Steel Pipe	150 mm, all ages	12.4	.5616	1.69
	200 mm, all ages	0	.1184	1.94
	250 mm, all ages	27.3	.7483	2.13
PVC Pipe	150 mm, 0-30 yrs	0	.0000	0

a. Values shown increased by a factor of 10^6

Table 6.6: Estimation of Parameter K1

the independent breaks are clustered on certain days can be obtained by adding a constant, K2, to Eq. 6.46.

$$P(I_1) = K2(LBIx)e^{LBIx} \tag{6.47}$$

where

- $K2$ is a constant used to measure the extent of concentration of independent breaks on particular days

If the probability of occurrence of pipe breaks is equal for all days of the year K2 can be expected to be close to one. On the other hand if there is a contributing cause which makes pipe breaks more likely to occur on certain days of the year K2 should be less than one. K2 is estimated by performing linear regression of P(I1) on $LBIxe^{-LBIx}$. The estimated values of K2 are shown in Table 6.7. For the 200 mm and 250 mm ductile steel and the 150 mm PVC pipe categories the r^2 values are too small to be statistically significant. For these categories it was assumed that $K2 = 1$.

6.5.5 Comparison of predicted to sampled probabilities

Once all four parameters have been estimated for all pipe categories, the estimated parameter values are substituted into Eqs. 6.12, 6.17, 6.27, and 6.37 to generate the probability distributions for days with m = 0, 1, 2, and 3 breaks. The predicted probabilities are compared with sampled frequencies for days with exactly 0, 1, 2, and 3 breaks. The comparison for days with no breaks is shown in Figs. B.28-B.46. For days with exactly 1 break the comparison is shown in Figs. B.47-B.65. For days with exactly two breaks the results are shown in Figs. B.66-B.84. The days with exactly three breaks the results are shown in Figs. B.85-B.103. While not perfect the agreement is still quite close for all cases.

Pipe Category Pipe material, diameter, age (mm) (yr)		Regression Parameters K2	r^2	Standard Deviation of $K2$
Cast Iron Pipe	150 mm, 0-19 yrs	.454	.7660	.0000195
	150 mm, 20-30 yrs	.546	.8829	0.0000186
	150 mm, over 30 yrs	.382	.4619	0.0000402
	200 mm, 0-30 yrs	.736	.7954	.0000214
	200 mm, over 30 yrs	.454	.6628	.0000255
	250 mm, 0-30 yrs	.613	.7945	.0000202
	250 mm, over 30 yrs	.490	.6905	.0000261
	300 mm, 0-30 yrs	.362	.5843	.0000165
	300 mm, over 30 yrs	.183	.4178	.000779
Asbestos Cement Pipe	150 mm, 0-30 yrs	.763	.5427	0.00000501
	150 mm, over 30 yrs	.690	.4667	0.0000091
	200 mm, 0-30 yrs	.565	.4273	0.0000126
	200 mm, over 30 yrs	.236	.4710	0.0000454
	250 mm, all ages	.420	.3306	0.0000105
	300 mm, all ages	.753	.6508	0.0000057
Ductile Steel Pipe	150 mm, all ages	.825	.8619	.0000045
	200 mm, all ages	1.000	.2847	.0000062
	250 mm, all ages	1.000	.1766	.0000045
PVC Pipe	150 mm, 0-30 yrs	1.000	.1571	0.0000011

Table 6.7: Estimation of Parameter K2

6.6 Confidence Limits for the Parameters LBI, K0, K1, and K2

The purpose of estimating these confidence limits is to determine limits on the values of exactly m breaks in a day. To determine these bounds, confidence limits are calculated on LBI, $K0$, $K1$, and $K2$, for each pipe category. Bonferroni confidence limits are used because, for each pipe category, multiple inferences (for each category, between 1 and 4 parameters must be estimated) are made from the same data. It should be noted that these confidence limits also are based on the length of pipe in each category. In addition to the estimation of the probability that exactly m pipes fail in a day, recall that the use of Eq. 5.6 to estimate network reliability requires the estimation of the probability of network failure given that m links from a pipe category have failed simultaneously. This conditional failure probability is estimated by a single parameter $\hat{p}$ for each failure case. As a result, separate Bonferroni confidence intervals must be determined for each network considered. These confidence intervals are presented in Chapters 9 and 10 for each network considered.

6.7 Confidence Limits for Total Probabilities

As outlined in Section 5.2, confidence limits can be calculated for both the coverage and the cumulative probability of network failure. The Bonferroni confidence limits and Eqs. 5.9 - 5.11 can be used to calculate the variances of both the coverage and the cumulative probabilities of nework failure. The t-distribution can be used calculate confidence limits and for hypothesis testing.

6.8 Limitations of this Method

This method of estimating exactly m breaks in a day gives reasonable results. The concept of independent breaks that are distributed as a Poisson distribution gives quite good results, especially for cast iron pipe. The concept that dependent breaks which occur on the same and subsequent days follow the Poisson distribution is also well supported. The use of a constant parameter, K2, to take into account the clustering of failure events in time is recognised as being somewhat simplistic. This aspect is an area in which further research is needed.

Chapter 7

Description of Model Networks

In the previous chapters, a method to estimate network reliability was developed. In this chapter the method is applied to the network detailed in Walski et al. (1987). This network was chosen because it has become recognized in the literature as a test case for reliable water distribution networks and because several solutions are available in Walski et al. (1987). (The detailed specifications of each network solution are given in this chapter.) It should be noted that the solutions in Walski et al. (1987) were developed using heuristic methods to ensure sufficient network reliability. In this thesis, probability based reliability estimates are made for each network.

For convenience, the original network specifications are reproduced from Walski et al. (1987). The network layout is shown in Fig. 7.1 with the fixed pipe characteristics being shown in Table 7.1 and the node characteristics being shown in Table 7.2. Note that it has been assumed that all pipe in the network is cast iron because only cast iron pipe was used in the network in Walski et al. (1987). The variable pipe characteristics for the original network are shown in Table 7.3. Note that pipes 68 through 76 have not been specified because they were not part of the original network in Walski et al. (1987) but were links to be designed in developing the network solution. The daily demand pattern is shown in Table 7.4. The network performance criteria specified in

Walski et al. (1987) include the provision of sufficient flow and pressure for normal operating conditions, instantaneous peak demand, and fire flows. All criteria apply to both 1985 demands and projected demands for 2005. It should be noted that no duration is specified for either the instantaneous peak flows or the fire flows. Since the instantaneous peak flows and fire flows are not well defined, only normal operating conditions, with the projected demands in 2005, are considered in this thesis. More specifically, the demand levels from the highest daily demand period are used. It should be noted that this method of estimating reliability can be used for fire flows and instantaneous peak demand, by using the appropriate nodal demands

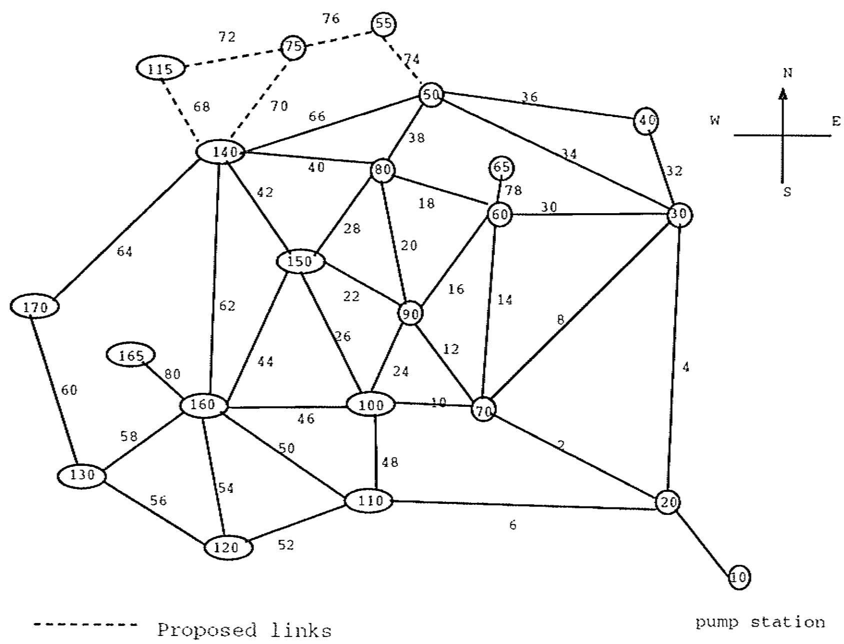


Figure 7.1: Original Network Layout from Walski, 1987

7.1 Assumed daily system operation plan

The daily demand pattern, shown in Table 7.4, divides the daily demand curve into eight time periods, each of three hours duration. Seven different demand levels rang-

Link Number	From Node	To Node	Link Length (feet)	Link Number	From Node	To Node	Link Length (feet)
1	10	20	100	42	140	150	6000
2	20	70	12000	44	150	160	6000
4	20	30	12000	46	100	160	6000
6	20	110	12000	48	100	110	6000
8	30	70	9000	50	110	160	6000
10	70	100	6000	52	110	120	6000
12	70	90	6000	54	120	160	9000
14	60	70	6000	56	120	130	6000
16	60	90	6000	58	130	160	6000
18	60	80	6000	60	130	170	6000
20	80	90	6000	62	140	160	6000
22	90	150	6000	64	140	170	12000
24	90	100	6000	66	50	140	12000
26	100	150	6000	68	115	140	6000
28	80	150	6000	70	75	140	6000
30	30	60	6000	72	75	115	6000
32	30	40	6000	74	50	55	6000
34	30	50	9000	76	55	75	6000
36	40	50	6000	78	60	65	100
38	50	80	6000	80	160	165	100
40	80	140	6000				

Table 7.1: Fixed data for links for all networks

Node	Elevation (ft)	Required Pressure (psi)	Required Flow (gpm)	Node	Elevation (ft)	Required Pressure (psi)	Required Flow (gpm)
10	280	pump		90	50	40	1000
20	20	40	500	100	50	40	500
30	50	40	200	110	50	40	500
40	50	40	200	115	80	40	600
50	50	40	600	120	120	40	400
55	80	40	600	130	120	40	400
60	50	40	500	140	80	40	400
65	225	tank	–	150	120	40	400
70	50	40	500	160	120	40	1000
75	80	40	600	165	225	tank	–
80	50	40	500	170	120	40	400

Table 7.2: Node characteristics table

Link Number	Link Diameter (inches)	Pipe Age (years)	Chezy's C	Link Number	Link Diameter (inches)	Pipe Age (years)	Chezy's C
1	20	50	120	42	8	50	120
2	16	50	70	44	8	50	120
4	12	50	120	46	8	50	120
6	12	50	70	48	8	50	120
8	12	50	70	50	10	50	120
10	12	50	70	52	8	50	120
12	10	50	70	54	8	50	120
14	12	50	70	56	8	50	120
16	10	50	70	58	10	50	120
18	12	50	70	60	8	50	120
20	10	50	70	62	8	50	120
22	10	50	70	64	8	50	120
24	10	50	70	66	8	50	120
26	12	50	70	68			
28	10	50	70	70			
30	10	50	120	72			
32	10	50	120	74			
34	10	50	120	76			
36	10	50	120	78	12	50	120
38	10	50	120	80	12	20	120
40	10	50	120				

Table 7.3: Link data for original network

Time Period	Demand as a Proportion of Average Daily Flow
24:00 - 03:00	0.6
03:00 - 06:00	0.7
06:00 - 09:00	1.1
09:00 - 12:00	1.2
12:00 - 15:00	1.3
15:00 - 18:00	1.2
18:00 - 21:00	1.0
21:00 - 24:00	0.9

Table 7.4: Daily demand pattern (average = 1.0)

ing from 60% of average daily demand to 130% of average daily demand are identified in this table. The nodal demands, for each demand level, are determined by multiplying the average demand, given in Table 7.2, by the appropriate factor, from Table 7.4.

Demands vary significantly during each day. The variation in demand levels has a significant impact on system operation. During periods of low demand, surplus hydraulic capacity can be used to fill the system storage tanks. The water stored in the storage tanks can then be used to meet flow and pressure requirements at peak demand levels.

This use of storage facilities complicates the task of simulating network performance and greatly increases the computational burden associated with the simulation. The approach to network simulation mentioned in Walski et al. (1987), involves the simulation of the network over a 24 hour period. The time interval used for the simulation is not given, but because of the highly non-linear relationship, between withdrawals from the storage tanks and flows in the network, the use of a small time interval seems reasonable. The use of a small time interval, however, implies a large increase in the computational burden associated with the calculation. For example, the use of a 15 minute time interval rather than a single iteration per day increases the number of times that the network must be solved, for each simulation by a factor of 96. As a result of these considerations, the following simplified procedure, using the daily demand curve and total storage capacity, is developed.

It should be noted that the storage capacity is limited. The storage tanks have a capacity of 250,000 gallons (US) for the base case. Walski et al. (1987) specify that fire flows must be met with the tanks 25% full. This requirement is taken to mean that the system is run with a reserve of 25% or 62,500 gallons (US) in each tank. If no additional storage is provided, the remaining 187,500 gallons U.S. storage can provide a maximum withdrawal of 1041.7 gallons per minute (gpm) for the three hour peak demand period. This withdrawal level has been reduced to 800 gpm in order to maintain storage capacity for withdrawals during the periods with elevated, but less than peak demands of 1.1 and 1.2 times average daily demand. It was assumed that withdrawals from storage are required in those periods when demand is greater than the average daily demand.

However, it is assumed that the peak demand period is critical, and hence it is the only case considered during the Monte Carlo simulation of network performance.

7.2 Example Networks

Four solution networks are considered. These solutions are:

1. non-optimal system expansion

2. the solution proposed by Gessler in Walski et al. (1987)

3. the solution proposed by Lee et al. in Walski et al. (1987)

4. the solution proposed by Morgan and Goulter in Walski et al. (1987)

7.3 'Non-Optimal' system expansion specification

The first network considered is a system expanded on a trial and error basis. A network solver was used to check the hydraulic adequacy of the network. It is assumed that pipes to the south and east of link 28 (specified as the older part of the city in Walski et al., 1987) are replaced. This assumption was made because most of these pipes would be over 70 years old in the year 2005.

The network layout for this non-optimal system expansion is shown in Fig. 7.2. The specifications of the new and replaced pipes in this network are given in Table 7.7 while the specifications of the tanks and pumps in this network for all demand levels are given in Table 7.8. Nodal pressures and flows were checked against required pressures and flows, for all demand levels, and the corresponding withdrawal (fill) levels are given in Table 7.8. Nodal pressures and flows were found to be adequate. It can thus be seen that this network has sufficient hydraulic capacity if no links have been broken.

Link Number	Link Diameter (inches)	Pipe Age (years)	Chezy's C	Link Number	Link Diameter (inches)	Pipe Age (years)	Chezy's C
2	20	20	130	22	12	20	130
4	18	20	130	24	10	20	130
6	18	20	130	26	12	20	130
8	12	20	130	28	10	20	130
10	16	20	130	68	10	20	130
12	12	20	130	70	8	20	130
14	12	20	130	72	8	20	130
16	10	20	130	74	8	20	130
18	12	20	130	76	10	20	130
20	10	20	130				

Table 7.5: Data for changed links in the non-optimal network

	Node Number		
	10	65	165
	Node Type		
Demand Level	Pump Head (ft)	Tank Withdrawals (Fill)(gpm)	Tank Withdrawals (Fill)(gpm)
.6	280	(800)	(800)
.7	280	(800)	(800)
.9	280	(100)	(100)
1.0	280	0	0
1.1	280	100	100
1.2	280	100	100
1.3	300	800	800

Table 7.6: Specification of tanks and pumps - non-optimal network

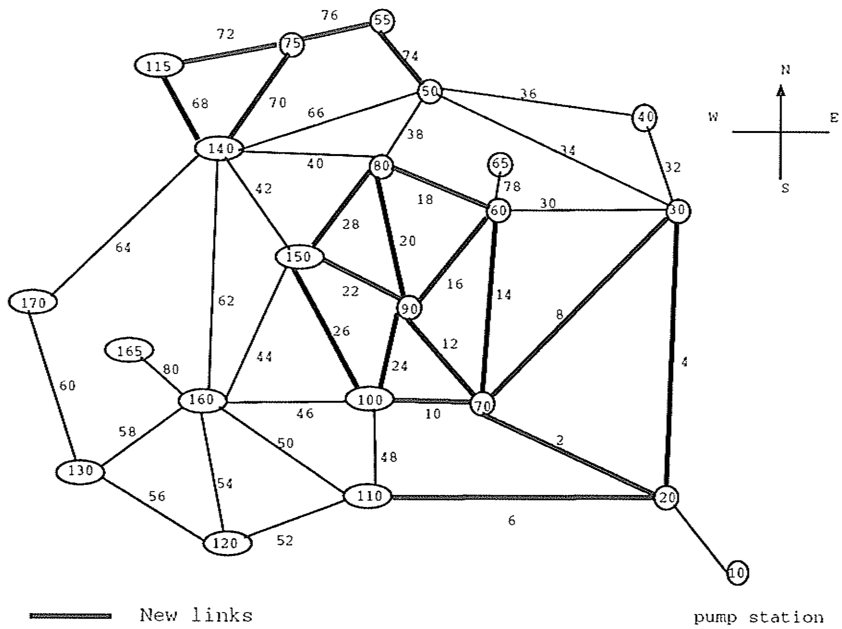


Figure 7.2: Non-Optimal Network Layout

7.4 The Gessler network specification

The third network considered is a system expansion proposed by Gessler in Walski et al. (1987). A network solver was again used to check the hydraulic adequacy of the network. This network is not hydraulically sufficient, if flows are limited to prevent any of the storage tanks from being emptied. However, as the network was quite close to hydraulic sufficiency, 'increased' flows were assumed to provide the required hydraulic capacity. These increased flows are equivalent to the assumption that the amount of overhead storage is greater than that specified by Gessler. The larger sized tank assumed, for this network, at node 150 makes significantly larger withdrawals from the tank possible. While the specification of a larger sized tank does not violate any of the network requirements as specified in Walski et al. (1987), the resulting increased cost of the network was not reflected in Gessler's solution. In addition, it should be noted that the design of a tank with insufficient capacity can be a serious error because after a tank has been built increasing its capacity can be quite difficult.

The network layout for this case is shown in Fig. 7.3. Note that an extra node, 175, and an extra link, number 82, from node 175 to node 150, have been added to the network, to take into account the addition of an overhead tank, connected to node 150. The specifications of the pipes, in this network, are given in Table 7.9. The specifications of the tanks and pumps in this network for all demand levels are given in Table 7.10. Nodal pressures and flows were checked against required pressures and flows for all demand levels and the corresponding withdrawal (fill) levels are given in Table 7.10. Nodal pressures and flows were found to be adequate. It can thus be seen that the Gessler solution, with the assumed extra storage, has sufficient hydraulic capacity if no links have been broken.

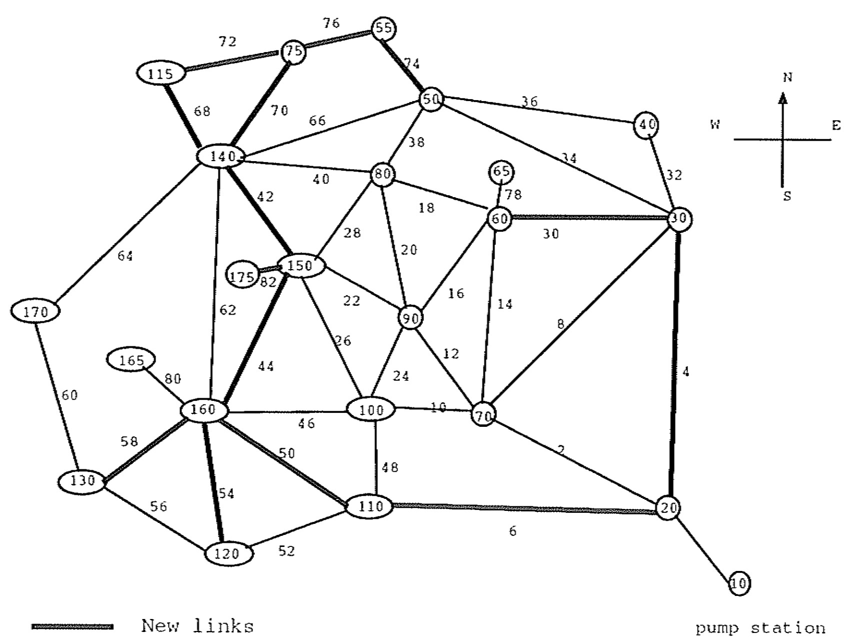


Figure 7.3: Gessler Network Layout

Link Number	Link Diameter (inches)	Pipe Age (years)	Chezy's C	Link Number	Link Diameter (inches)	Pipe Age (years)	Chezy's C
4	14	20	130	60	14	20	130
6	14	20	130	68	12	20	130
30	10	20	130	70	12	20	130
42	14	20	130	72	6	15	130
44	18	20	130	74	14	20	130
50	24	20	130	76	6	15	130
54	8	20	130	82	12	20	130
58	16	20	130				

Table 7.7: Data for changed links in the Gessler network

7.5 Lee et al. solution specifications

The fourth network is a system expansion proposed by Lee et al. in Walski et al. (1987). A network solver was again used to check the hydraulic adequacy of the network.

The network layout is shown in Fig. 7.4. Note that an extra node, 175, and an extra link, 175-140, have been added to the network to take into account the specification of an additional overhead tank connected to node 140. The specifications of the pipes added to or changed in this network are given in Table 7.11. The specifications of the tanks and pumps in this network for all demand levels are given in Table 7.12. It should be noted that the tank proposed by Lee et al. at node 140 is 100,000 gallons smaller than that proposed by Gessler. The maximum withdrawal rate from the tank is thus smaller than in the solution proposed by Gessler and is listed in Table 7.12. Nodal pressures and flows were again checked against required pressures and flows,

Demand Level as a Proportion of Average Daily Demand	Node Number			
	10	65	175	165
	Node Type			
	Pump Head (ft)	Tank Withdrawals (Fill)(gpm)	Tank Withdrawals (Fill)(gpm)	Tank Withdrawals (Fill)(gpm)
.6	280	(800)	(2200)	(800)
.7	280	(800)	(2200)	(800)
.9	280	(100)	(400)	(100)
1.0	280	0	0	0
1.1	280	100	450	100
1.2	280	100	450	100
1.3	310	1200	4400	1350

Table 7.8: Specification of tanks and pumps - Gessler network

for all demand levels for this network and the corresponding withdrawal (fill) levels are given in Table 7.12. Nodal pressures and flows were found to be adequate. It can, thus, be seen that this network also has sufficient hydraulic capacity if no links have been broken.

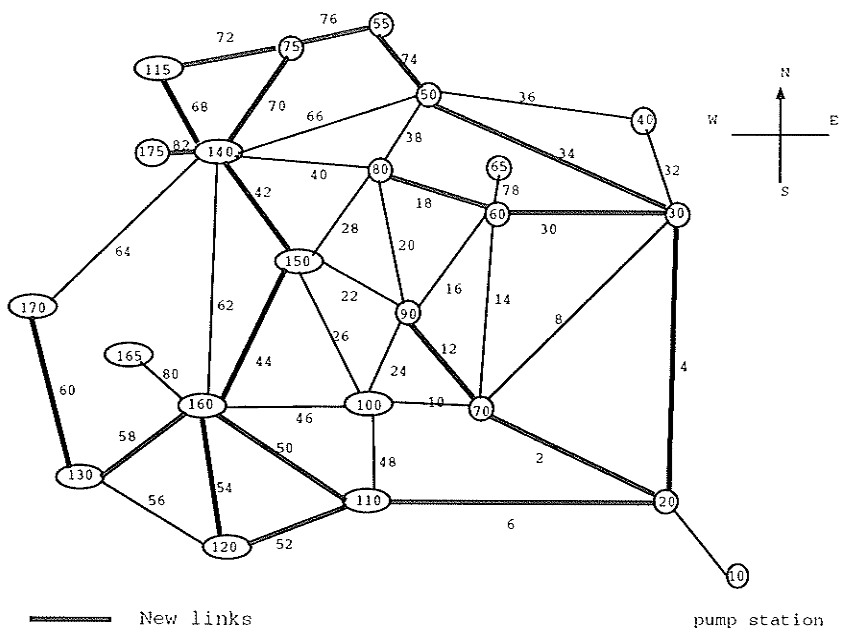


Figure 7.4: Lee et al. Network Layout

7.6 Morgan and Goulter network specifications

The last network considered is a system expansion proposed by Morgan and Goulter in Walski (1987) The network layout is shown in Fig. 7.5. Note that extra nodes, 175 and 185, and extra links, 175-75 and 185-170, have been added to the network to take into account the provision of additional overhead tanks. The specifications of the pipes in this network are given in Table 7.13 while the specifications of the tanks and pumps in this network for all demand levels, are given in Table 7.14. These withdrawal values, corresponding to the sizes of these tanks, are listed in Table 7.14. A network

Link Number	Link Diameter (inches)	Pipe Age (years)	Chezy's C	Link Number	Link Diameter (inches)	Pipe Age (years)	Chezy's C
2	17	20	130	52	12	20	130
4	15	20	130	54	6	20	130
6	19	20	130	58	12	20	130
12	15	20	130	60	12	20	130
18	8	20	130	68	12	20	130
30	10	20	130	70	12	20	130
34	14	20	130	72	6	15	130
42	14	20	130	74	12	20	130
44	18	20	130	76	6	15	130
50	12	20	130	80	12	20	130

Table 7.9: Data for changed links in the Lee et al. network

Demand Level as a Proportion of Average Daily Demand	Node Number			
	10	65	175	165
	Node Type			
	Pump Head (ft)	Tank Withdrawals (Fill)(gpm)	Tank Withdrawals (Fill)(gpm)	Tank Withdrawals (Fill)(gpm)
.6	280	(800)	(2000)	(800)
.7	280	(800)	(2000)	(800)
.9	280	(100)	(400)	(100)
1.0	280	0	0	0
1.1	280	100	450	100
1.2	280	100	450	100
1.3	300	800	2600	800

Table 7.10: Specification of tanks and pumps - Lee et al. network

solver was used to check the hydraulic adequacy of the network. Nodal pressures and flows obtained by using the network solver were checked against required pressures and flows for all demand levels and the corresponding withdrawal (fill) levels are given in Table 7.14. Nodal pressures and flows were found to be adequate. It can be seen therefore that this network also has sufficient hydraulic capacity if no links have been broken.

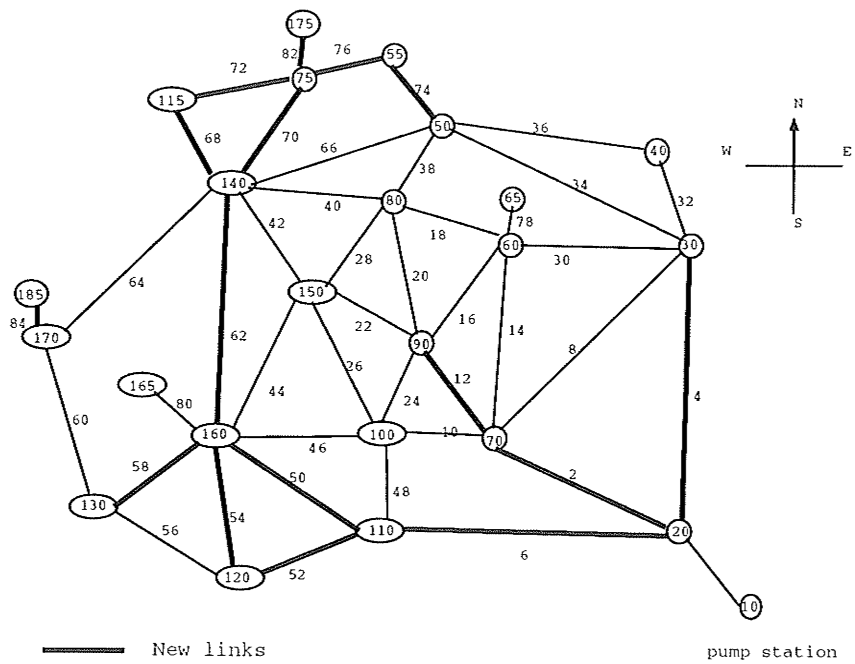


Figure 7.5: Morgan and Goulter Network Layout

7.7 Determining covering sets for pipe removal set probabilities

As discussed previously, in order to determine the reliability of a network, containing several different types of pipes, it is necessary to determine a covering set from among all possible combinations of pipe removal sets. The same procedure is used for all

Link Number	Link Diameter (inches)	Pipe Age (years)	Chezy's C	Link Number	Link Diameter (inches)	Pipe Age (years)	Chezy's C
2	12	20	130	68	10	20	130
4	8	20	130	70	10	20	130
6	20	20	130	72	8	20	130
12	10	20	130	74	8	20	130
50	18	20	130	76	10	20	130
52	6	15	130	80	12	20	130
54	6	15	130	82	12	20	130
58	10	50	130	84	12	20	130
62	12	20	130				

Table 7.11: Data for changed links in the Morgan and Goulter network

networks described in the previous section. The steps that are followed are:

1. Determine the length of pipe by pipe category.

2. Calculate the estimated probability of days with exactly 0, 1, 2, and 3 breaks. Bonferroni limits and expected values are calculated for these probabilities.

3. Choose pipe breakage events consisting of the removal of combinations of links from the various pipe categories. A sufficient number of combinations are chosen to ensure that almost all probable failure cases included.

The pipe length for each network is calculated using the pipe lengths given in Table 7.1. The pipe categories shown in Table 7.15 are determined using the original pipe characteristics data in Table 7.3 and the updated pipe characteristics in Tables 7.7, 7.9, 7.11, or 7.13 for the pipes in each proposed network. From the lengths of the

	Node Number			
	10	65	175	165
	Node Type			
Demand Level	Pump Head (ft)	Tank Withdrawals (Fill)(gpm)	Tank Withdrawals (Fill)(gpm)	Tank Withdrawals (Fill)(gpm)
.6	280	(800)	(2000)	(800)
.7	280	(800)	(2000)	(800)
.9	280	(100)	(400)	(100)
1.0	280	0	0	0
1.1	280	100	450	100
1.2	280	100	450	100
1.3	300	800	2600	800

Table 7.12: Specification of tanks and pumps - Morgan and Goulter network

Pipe Category	Pipe Category Description
6 0-19	Cast iron pipe, 6 in. diameter, 0-19 years old
8 0-30	Cast iron pipe, 8 in. diameter, 0-30 years old
8 over 30	Cast iron pipe, 8 in. diameter, more than 30 years old
10 0-30	Cast iron pipe, 10 in. diameter, 0-30 years old
10 over 30	Cast iron pipe, 10 in. diameter, more than 30 years old
12 0-30	Cast iron pipe, 12 in. diameter, 0-30 years old
12 over 30	Cast iron pipe, 12 in. diameter, more than 30 years old

Table 7.13: Pipe categories and abbreviations

individual pipes in the pipe category the total pipe length for each pipe category is then determined. The summarized pipe lengths for all categories are shown for each network in Tables 7.14, 7.15, 7.16 and 7.17.

The probability of days with exactly 0, 1, 2, and 3 breaks are then estimated using Eqs. 6.12, 6.17, 6.27, and 6.37. These equations require the pipe length in Tables 7.14 - 7.17. In addition, the use of these equations requires estimation of the four parameters LBI, $K0$, $K1$, and $K2$. The mean and standard error of these parameters LBI, $K0$, $K1$, and $K2$ based on the City of Winnipeg data are given in Tables 6.3 - 6.7. Bonferroni confidence limits are calculated for each of the parameters for each network because these estimates require the simultaneous estimation of parameters for all pipe categories and for all failure cases in the covering set. The limits vary from network to network because the length of pipe, the number of pipe categories, and the number of failure cases to form a covering set can vary. Upper and lower limits are calculated for the probability of days with exactly 0, 1, 2, and 3 failures for all networks. In addition, the expected values of the probabilities of days with exactly 0, 1, 2, and 3 failures are calculated for all networks. These probabilities are shown in Tables 7.18,7.19,7.20, and 7.21.

Pipe Category Pipe material, diameter, age	Length (ft)	Length (km)
Cast iron, 8 in, 0-30 years	18,000	5.49
8 in, over 30 years	33,000	10.06
10 in, 0-30 years	36,000	10.97
10 in, over 30 years	51,000	15.54
12 in, 0-30 year	81,100	24.69
12 in, over 30 years	48,200	14.69

Table 7.14: Pipe length by pipe category - non-optimal network

Pipe Category Pipe material, diameter, age	Length (ft)	Length (km)
Cast iron, 6 in, 0-19 years	12,000	3.66
8 in, 0-30 years	9,000	2.74
8 in, over 30 years	54,000	16.46
10 in, 0-30 years	6,000	1.82
10 in, over 30 years	69,000	21.03
12 in, 0-30 year	72,100	21.98
12 in, over 30 years	45,300	13.81

Table 7.15: Pipe length by pipe category - Gessler network

Pipe Category	Length	
Pipe material, diameter, age	(ft)	(km)
Cast iron, 6 in, 0-19 years	21,000	6.40
8 in, 0-30 years	6,000	1.83
8 in, over 30 years	48,000	14.63
10 in, 0-30 years	6,000	1.83
10 in, over 30 years	54,000	16.46
12 in, 0-30 year	105,200	33.12
12 in, over 30 years	27,200	7.29

Table 7.16: Pipe length by pipe category - Lee et al. network

Pipe Category	Length	
Pipe material, diameter, age	(ft)	(km)
Cast iron, 6 in, 0-19 years	15,000	4.57
8 in, 0-30 years	24,000	7.32
8 in, over 30 years	54,000	16.46
10 in, 0-30 years	24,000	7.315
10 in, over 30 years	75,000	22.86
12 in, 0-30 year	36,300	11.06
12 in, over 30 years	39,000	11.89

Table 7.17: Pipe length by pipe category - Morgan and Goulter network

Pipe Category Diameter Age (in) (yr)	Probability of a Day with Exactly m Breaks					
	m=0			m=1		
	Lower Limit	Upper Limit	Expected Value	Lower Limit	Upper Limit	Expected Value
8 0-30	.997126	.997227	.997177	.001768	.001822	.001795
8 over 30	.993892	.994132	.994012	.002406	.002495	.002451
10 0-30	.993907	.995634	.99477	.002314	.003035	.002702
10 over 30	.989282	.992066	.990673	.003492	.004537	.004038
12 0-30	.987443	.991978	.989708	.002749	.00417	.003484
12 over 30	.993239	.996449	.994842	-.0000084E	.0000208	.00000203

Pipe Category Diameter Age (in) (yr)	Probability of a Day with Exactly m Breaks					
	m=2			m=3		
	Lower Limit	Upper Limit	Expected Value	Lower Limit	Upper Limit	Expected Value
8 0-30	.000795	.000929	.000862	.000312	.000412	.000362
8 over 30	.001953	.002161	.002057	.001041	.001194	.001117
10 0-30	.001741	.002095	.001925	.000704	.000966	.00084
10 over 30	.002837	.003489	.003172	.001356	.00184	.001605
12 0-30	.002846	.004159	.003517	.001684	.002669	.002188
12 over 30	.001623	.002971	.00231	.001083	.002095	.001599

Table 7.18: Bonferroni confidence limits for component failures - non-optimal network

Pipe Category Diameter Age (in) (yr)	Probability of a Day with Exactly m Breaks					
	m=0			m=1		
	Lower Limit	Upper Limit	Expected Value	Lower Limit	Upper Limit	Expected Value
6 0-19	.998462	.998905	.998683	.000403	.00063	.000507
8 0-30	.998565	.998615	.99859	.000876	.00092	.000898
8 over 30	.990026	.990417	.990222	.003888	.004099	.003993
10 0-30	.998986	.999275	.99913	.000379	.000519	.000451
10 over 30	.985518	.989284	.987399	.004647	.006227	.005444
12 0-30	.988808	.992861	.990833	.002433	.003755	.003105
12 over 30	.993642	.996662	.995151	.000583	.001123	.000854
Pipe Category Diameter Age (in) (yr)	Probability of a Day with Exactly m Breaks					
	m=2			m=3		
	Lower Limit	Upper Limit	Expected Value	Lower Limit	Upper Limit	Expected Value
6 0-19	.00035	.000693	.000517	.000183	.000302	.000239
8 0-30	.000383	.000476	.00043	.000154	.000206	.00018
8 over 30	.003106	.003638	.003371	.001697	.001975	.001836
10 0-30	.000273	.000364	.000319	.000115	.000162	.000139
10 over 30	.003642	.004952	.004297	.001819	.002541	.002181
12 0-30	.002456	.003793	.003129	.001488	.002395	.001945
12 over 30	.001269	.002382	.001831	.000838	.001648	.001247

Table 7.19: Bonferroni confidence limits for component failures - Gessler network

Pipe Category Diameter Age (in) (yr)	Probability of a Day with Exactly m Breaks					
	m=0			m=1		
	Lower Limit	Upper Limit	Expected Value	Lower Limit	Upper Limit	Expected Value
6 0-19	.997312	.998086	.997699	.000704	.0011	.000886
8 0-30	.999041	.999075	.999058	.000585	.000615	.0006
8 over 30	.99113	.991478	.991304	.00346	.003648	.003553
10 0-30	.998986	.999275	.99913	.000379	.000519	.000451
10 over 30	.988647	.991603	.990124	.003647	.004891	.004274
12 0-30	.988793	.992851	.99082	.002437	.00376	.003109
12 over 30	.996178	.997995	.997086	.00035	.000676	.000514

Pipe Category Diameter Age (in) (yr)	Probability of a Day with Exactly m Breaks					
	m=2			m=3		
	Lower Limit	Upper Limit	Expected Value	Lower Limit	Upper Limit	Expected Value
6 0-19	.000613	.001211	.000904	.00032	.000528	.000418
8 0-30	.000256	.000318	.000287	.000103	.000138	.00012
8 over 30	.00276	.003231	.002995	.001507	.001754	.00163
10 0-30	.000273	.000364	.000319	.000115	.000162	.000139
10 over 30	.002849	.00387	.00336	.00142	.001981	.001701
12 0-30	.002459	.003798	.003134	.00149	.002398	.001948
12 over 30	.000762	.001428	.001099	.000502	.000987	.000748

Table 7.20: Bonferroni confidence limits for component failures - Lee et al. network

Pipe Category Diameter Age (in) (yr)	Probability of a Day with Exactly m Breaks					
	m=0			m=1		
	Lower Limit	Upper Limit	Expected Value	Lower Limit	Upper Limit	Expected Value
6 0-19	.99808	.998633	.998356	.000503	.000786	.000633
8 0-30	.99617	.996304	.996237	.002333	.00245	.002391
8 over 30	.990026	.990417	.990222	.003888	.004099	.003993
10 0-30	.995931	.997088	.99651	.001521	.002077	.001806
10 over 30	.984268	.988356	.98631	.005046	.00676	.00591
12 0-30	.994353	.996401	.995377	.001229	.001901	.00157
12 over 30	.994523	.997125	.995823	.000502	.000968	.000736

Pipe Category Diameter Age (in) (yr)	Probability of a Day with Exactly m Breaks					
	m=2			m=3		
	Lower Limit	Upper Limit	Expected Value	Lower Limit	Upper Limit	Expected Value
6 0-19	.000437	.000865	.000646	.000228	.000377	.000298
8 0-30	.001026	.001275	.00115	.000414	.000553	.000483
8 over 30	.003106	.003638	.003371	.001697	.001975	.001836
10 0-30	.001098	.001465	.001283	.000464	.000652	.000559
10 over 30	.003959	.005386	.004673	.001979	.002767	.002374
12 0-30	.001234	.001902	.001571	.000746	.001198	.000974
12 over 30	.001092	.00205	.001576	.000721	.001418	.001073

Table 7.21: Bonferroni confidence limits for component failures - Morgan and Goulter network

A covering set, for estimation of pipe failure probabilities, can then be determined, as outlined in Chapter 5. A covering set was chosen for each of the networks. The same procedure was used for all four networks to choose failure events to be included in the covering set. The first failure cases chosen included events comprised of all days with 0 and/or 1 break. Next days with exactly 2 breaks were chosen. Finally, days with exactly 3 breaks were chosen. The covering set for the non-optimal network is shown in Tables 7.22, 7.23, and 7.24. The cumulative probability of all failure cases included in the covering set for the non-optimal network was calculated after each step by summing the expected value of the probability of occurrence over all failure cases. For example the cumulative probability covered in Table 7.23 includes the sum of expected values for failure cases 1 through 12 and the 'no break' case. Days with exactly one break from two different pipe categories were examined, but not included in the covering set because under the assumption of independence of failure events involving different types of pipe the probability of such events was several orders of magnitude less than the removal events considered. The covering set for the Gessler network is shown in Tables 7.25, 7.26, and 7.27. The covering set for the Lee et al. network is shown in Tables 7.28, 7.29, and 7.30. The covering set for the Morgan and Goulter network is shown in Tables 7.31, 7.32, and 7.33.

It is interesting to note that that the expected cumulative probabilities (shown in Tables 7.24, 7.27, 7.30, and 7.33) are very close to 1 for all four networks showing that the covering sets do include almost all probable component failure events. The variance of the cumulative probabilities found using Eq. 5.9, was calculated for all four networks and confidence limits were also determined. The standard deviation and confidence intervals are shown for all four networks in Table 7.34. The lower bounds of these confidence intervals show that over 99.0% of component failure probabilities for each network are accounted for by the respective covering sets. From Eq. 5.7 the

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
no break		.955699	.967912	.961786
1	8 0-30	.001695	.001769	.001732
2	8 over 30	.002314	.002429	.002371
3	10 0-30	.002225	.002951	.002613
4	10 over 30	.003373	.004426	.00392
5	12 0-30	.002661	.004069	.003385
6	12 over 30	-.0000081	.0000202	.00000197
Cumulative probability covered		.		.9758089

Table 7.22: Covering set - days with exactly 0 or 1 break: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
7	8 0-30, 8 0-30	.000762	.000902	.000831
8	8 over 30, 8 over 30	.001878	.002104	.00199
9	10 0-30, 10 0-30	.001674	.002037	.001861
10	10 over 30, 10 over 30	.00274	.003404	.003079
11	12 0-30, 12 0-30	.002754	.004058	.003418
12	12 over 30, 12 over 30	.001562	.002885	.002234
Cumulative probability covered				.989222

Table 7.23: Covering set - days with exactly 2 breaks: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
11	8 0-30, 8 0-30, 8 0-30	.000299	.0004	.000349
12	8 over 30, 8 over 30, 8 over 30	.001001	.001162	.001081
13	10 0-30, 10 0-30, 10 0-30	.000677	.000939	.000812
14	10 over 30, 10 over 30, 10 over 30	.00131	.001795	.001558
15	12 0-30, 12 0-30, 12 0-30	.00163	.002604	.002126
16	12 over 30, 12 over 30, 12 over 30	.001042	.002035	.001546
Cumulative probability covered				.996695

Table 7.24: Covering set - days with exactly 3 breaks: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
no break		.954817	.966457	.960619
1	6 0-19	.000385	.000609	.000488
2	8 0-30	.000837	.00089	.000863
3	8 over 30	.00375	.004	.003873
4	10 0-30	.000363	.000502	.000433
5	10 over 30	.004502	.006084	.005296
6	12 0-30	.00235	.003655	.00301
7	12 over 30	.00056	.001089	.000824
Cumulative probability covered				.975406

Table 7.25: Covering set - days with exactly 0 or 1 break: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
8	6 0-19, 6 0-19	.000335	.00067	.000497
9	8 0-30, 8 0-30	.000367	.000461	.000413
10	8 over 30, 8 over 30	.002996	.00355	.00327
11	10 0-30, 10 0-30	.000261	.000352	.000307
12	10 over 30, 10 over 30	.003528	.004837	.004181
13	12 0-30, 12 0-30	.002372	.003692	.003034
14	12 over 30, 12 over 30	.001213	.002304	.001761
Cumulative probability covered				.988869

Table 7.26: Covering set - days with exactly 2 breaks: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
15	6 0-19, 6 0-19, 6 0-19	.000175	.000292	.00023
16	8 0-30, 8 0-30, 8 0-30	.000148	.0002	.000173
17	8 over 30, 8 over 30, 8 over 30	.001637	.001927	.001781
18	10 0-30, 10 0-30, 10 0-30	.00011	.000156	.000133
19	10 over 30, 10 over 30, 10 over 30	.001762	.002483	.002122
20	12 0-30, 12 0-30, 12 0-30	.001437	.002331	.001886
21	12 over 30, 12 over 30, 12 over 30	.000805	.001598	.001204
Cumulative probability covered				.996401

Table 7.27: Covering set - days with exactly 3 breaks: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
no break		.960701	.970698	.965686
1	6 0-19	.000678	.00107	.000857
2	8 0-30	.000563	.000597	.00058
3	8 over 30	.003353	.003572	.003461
4	10 0-30	.000365	.000504	.000436
5	10 over 30	.003544	.004788	.004169
6	12 0-30	.002367	.003676	.00303
7	12 over 30	.000338	.000658	.000497
Cumulative probability covered				.978716

Table 7.28: Covering set - days with exactly 0 or 1 break: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
8	6 0-19, 6 0-19	.00059	.001178	.000875
9	8 0-30, 8 0-30	.000246	.000309	.000277
10	8 over 30, 8 over 30	.002675	.003164	.002918
11	10 0-30, 10 0-30	.000263	.000353	.000308
12	10 over 30, 10 over 30	.002768	.003789	.003277
13	12 0-30, 12 0-30	.002389	.003713	.003054
14	12 over 30, 12 over 30	.000734	.001389	.001064
Cumulative probability covered				.990489

Table 7.29: Covering set - days with exactly 2 breaks: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
15	6 0-19, 6 0-19, 6 0-19	.000308	.000514	.000405
16	8 0-30, 8 0-30, 8 0-30	.000099	.000134	.000116
17	8 over 30, 8 over 30, 8 over 30	.00146	.001717	.001588
18	10 0-30, 10 0-30, 10 0-30	.000111	.000157	.000134
19	10 over 30, 10 over 30, 10 over 30	.00138	.00194	.001659
20	12 0-30, 12 0-30, 12 0-30	.001448	.002345	.001899
21	12 over 30, 12 over 30, 12 over 30	.000484	.00096	.000724
Cumulative probability covered				.997015

Table 7.30: Covering set - days with exactly 3 breaks Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
no break		.954209	.964822	.959501
1	6 0-19	.000481	.000759	.000608
2	8 0-30	.002235	.002373	.002303
3	8 over 30	.003747	.003993	.003869
4	10 0-30	.001457	.00201	.001739
5	10 over 30	.004892	.006599	.00575
6	12 0-30	.00118	.001841	.001514
7	12 over 30	.000481	.000937	.000709
Cumulative probability covered				.975993

Table 7.31: Covering set - days with exactly 0 or 1 break: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
8	6 0-19, 6 0-19	.000418	.000836	.00062
9	8 0-30, 8 0-30	.000982	.001234	.001108
10	8 over 30, 8 over 30	.002994	.003544	.003266
11	10 0-30, 10 0-30	.001052	.001417	.001235
12	10 over 30, 10 over 30	.003838	.005257	.004546
13	12 0-30, 12 0-30	.001185	.001842	.001514
14	12 over 30, 12 over 30	.001042	.001978	.001512
Cumulative probability covered				.989794

Table 7.32: Covering set - days with exactly 2 breaks: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Probability of Occurrence		
		Lower Limit	Upper Limit	Expected Value
15	6 0-19, 6 0-19, 6 0-19	.000218	.000364	.000287
16	8 0-30, 8 0-30, 8 0-30	.000397	.000535	.000466
17	8 over 30, 8 over 30, 8 over 30	.001636	.001924	.001779
18	10 0-30, 10 0-30, 10 0-30	.000445	.000631	.000538
19	10 over 30, 10 over 30, 10 over 30	.001919	.002701	.002309
20	12 0-30, 12 0-30, 12 0-30	.000716	.00116	.000939
21	12 over 30, 12 over 30, 12 over 30	.000692	.001372	.001034
Cumulative probability covered				.997147

Table 7.33: Covering set - days with exactly 3 breaks Morgan and Goulter network

Network	Standard Deviation	Lower Limit	Upper Limit	Maximum Uncounted Failure Probability
Non-optimal	.002447	.990390	1.002999	.00961
Gessler	.002355	.990334	1.002468	.009666
Lee et al.	.002025	.991800	1.00223	.008200
Morgan and Goulter	.002145	.991622	1.002671	.008378

Table 7.34: Confidence limits for cumulative probabilities - all networks

maximum uncounted failure probability for each network, found by subtracting the lower limit from one, is an upper bound on the probability of system failure.

In Chapter 8, Monte Carlo sampling from these covering sets is used to determine network reliability, based on connectivity for all four networks while in Chapter 9, Monte Carlo sampling is used to determine network reliability, based on adequacy of flow and pressure at all nodes.

Chapter 8

Estimation of Reliability Performance of Networks Using the Connectivity Criterion

In this chapter, the performance of the test networks described in Chapter 7 in terms of connectivity is discussed. When using the connectivity criterion a node is assumed to be connected, if an unbroken path exists from the node, to the source node. For the purposes of this study, it was assumed that any node disconnected, from the source pumps at node 10 was disconnected from the network.

Monte Carlo simulation is used to estimate the probability that the network becomes disconnected given that links have been removed. The simulation program works as follows:

1. A failure case which specifies the pipe categories in which pipe breaks are to occur is selected.

2. A pipe is selected randomly for each pipe category specified in the failure case. In this process the pipe selections are weighted by the length of the pipe with

respect to the total length of pipe, in the pipe category. For example if a pipe category contains 10 km of pipe and a particular pipe has a length of 1 km then that pipe would be chosen with a probability of .10.

3. The selected pipes are removed from the network.

4. The remaining network is checked for disconnected nodes.

5. The proportion of removal sets which cause disconnection is estimated by repeated trials. The error in the estimate is determined then using Eq. 4.22.

6. Trials are repeated until the error is reduced to a preset level.

The probability,of one or more nodes being disconnected will be calculated for each of the four enhanced networks discussed in Chapter 7. Confidence limits are developed for each network.

8.1 Estimating the Probability of Node Disconnection

The Bonferroni confidence limits for the probability of occurrence were presented in Tables 7.22, 7.23, and 7.24 for each case in the covering set of the non-optimal network. The estimated probability of failure as given by Eq. 4.18, given that a removal event has occurred in the non-optimal network, is shown in Tables 8.1, 8.2, and 8.3. Three tables are used for each covering set to match Tables 7.22, 7.23, and 7.24. Of particular interest in Tables 8.1, 8.2, and 8.3 are the categories for which the conditional probability of failure is .00. For these categories none of the combinations selected during the Monte Carlo simulation caused system failure. This result should not be surprising, especially in the cases in which only a single link is removed from

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
no break		.00	.00
1	8 0-30	.00	.00
2	8 over 30	.00	.00
3	10 0-30	.00	.00
4	10 over 30	.00	.00
5	12 0-30	.088916	.037738
6	12 over 30	.185512	.119848

Table 8.1: Conditional failure probabilities - days with exactly 0 or 1 break: non-optimal network

the network because one the design criteria given in Walski et al. (1988) is that the removal of a single link from the network does not cause system failure.

The upper and lower bounds for the probability of failure, defined as the occurrence of node disconnection, for each case are calculated using Eqs. 5.9 and 5.10. Upper bounds, for estimated error, with respect to these estimates is calculated, using Eq. 5.12. These results are shown in Tables 8.4, 8.5, and 8.6 which also contain the expected system failure probabilities.

A similar approach is used to calculate the estimated probability of network failure for the other three networks. The results for the Gessler, Lee et al., and Morgan and Goulter networks are presented in Tables 8.7-8.12, Tables 8.13-8.18, and Tables 8.19-8.24 respectively.

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
7	8 0-30, 8 0-30	.00	.00
8	8 over 30, 8 over 30	.00	.00
9	10 0-30, 10 0-30	.00	.00
10	10 over 30, 10 over 30	.00	.00
11	12 0-30, 12 0-30	.171472	.039141
12	12 over 30, 12 over 30	.343602	.119909

Table 8.2: Conditional failure probabilities - days with exactly 2 breaks: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
13	8 0-30, 8 0-30, 8 0-30	.094252	.029973
14	8 over 30, 8 over 30, 8 over 30	.00	.00
15	10 0-30, 10 0-30, 10 0-30	.00	.00
16	10 over 30, 10 over 30, 10 over 30	.00	.00
17	12 0-30, 12 0-30, 12 0-30	.250134	.043668
18	12 over 30, 12 over 30, 12 over 30	.492505	.119994

Table 8.3: Conditional failure probabilities - days with exactly 3 breaks: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
no break		.00	.00	.00
1	8 0-30	.00	.00	.00
2	8 over 30	.00	.00	.00
3	10 0-30	.00	.00	.00
4	10 over 30	.00	.00	.00
5	12 0-30	.000188	.000437	.000301
6	12 over 30	-.000001	.00000494	.000000365

Table 8.4: System failure probabilities - days with exactly 0 or 1 break: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
7	8 0-30, 8 0-30	.00	.00	.00
8	8 over 30, 8 over 30	.00	.00	.00
9	10 0-30, 10 0-30	.00	.00	.00
10	10 over 30, 10 over 30	.00	.00	.00
11	12 0-30, 12 0-30	.000442	.000773	.000586
12	12 over 30, 12 over 30	.000445	.001160	.000768

Table 8.5: System failure probabilities - days with exactly 2 breaks: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
11	8 0-30, 8 0-30, 8 0-30	.0000238	.0000435	.0000329
12	8 over 30, 8 over 30, 8 over 30	.00	.00	.00
13	10 0-30, 10 0-30, 10 0-30	.00	.00	.00
14	10 over 30, 10 over 30, 10 over 30	.00	.00	.00
15	12 0-30, 12 0-30, 12 0-30	.000373	.000707	.000532
16	12 over 30, 12 over 30, 12 over 3 0	.000452	.001121	.000761

Table 8.6: System failure probabilities - days with exactly 3 breaks: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
no break		.00	.00
1 2	8 0-30	.00	.00
3	8 over 30	.00	.00
4	10 0-30	.00	.00
5	10 over 30	.00	.00
6	12 0-30	.00	.00
7	12 over 30	.179577	.096458

Table 8.7: Conditional failure probabilities - days with exactly 0 or 1 break: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
8	6 0-19, 6 0-19	.00	.00
9	8 0-30, 8 0-30	.00	.00
10	8 over 30, 8 over 30	.00	.00
11	10 0-30, 10 0-30	.00	.00
12	10 over 30, 10 over 30	.014184	.028403
13	12 0-30, 12 0-30	.00	.00
14	12 over 30, 12 over 30	.320639	.119994

Table 8.8: Conditional failure probabilities - days with exactly 2 breaks: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
15	6 0-19, 6 0-19, 6 0-19	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.00	.00
18	10 0-30, 10 0-30, 10 0-30	.00	.00
19	10 over 30, 10 over 30, 10 over 30	.035408	.044404
20	12 0-30, 12 0-30, 12 0-30	.00	.00
21	12 over 30, 12 over 30, 12 over 30	.452484	.119979

Table 8.9: Conditional failure probabilities - days with exactly 3 breaks: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
no break		.00	.00	.00
1 2	8 0-30	.00	.00	.00
3	8 over 30	.00	.00	.00
4	10 0-30	.00	.00	.00
5	10 over 30	.00	.00	.00
6	12 0-30	.00	.00	.00
7	12 over 30	.0000742	.000247	.000148

Table 8.10: System failure probabilities - days with exactly 0 or 1 break: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
8	6 0-19, 6 0-19	.00	.00	.00
9	8 0-30, 8 0-30	.00	.00	.00
10	8 over 30, 8 over 30	.00	.00	.00
11	10 0-30, 10 0-30	.00	.00	.00
12	10 over 30, 10 over 30	.00000111	.000136	.0000593
13	12 0-30, 12 0-30	.00	.00	.00
14	12 over 30, 12 over 30	.000318	.000874	.000565

Table 8.11: System failure probabilities - days with exactly 2 breaks: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
15	6 0-19, 6 0-19, 6 0-19	.00	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.00	.00	
18	10 0-30, 10 0-30, 10 0-30	.00	.00	
19	10 over 30, 10 over 30, 10 over 30	.0000242	.000142	.0000751
20	12 0-30, 12 0-30, 12 0-30	.00	.00	.00
21	12 over 30, 12 over 30, 12 over 3 0	.000318	.000817	.000545

Table 8.12: System failure probabilities - days with exactly 3 breaks: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
no break		.00	.00
1	6 0-19	.00	.00
2	8 0-30	.00	.00
3	8 over 30	.00	.00
4	10 0-30	.00	.00
5	10 over 30	.00	.00
6	12 0-30	.156391	.106057
7	12 over 30	.00	.00

Table 8.13: Conditional failure probabilities - days with exactly 0 or 1 break: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
8	6 0-19, 6 0-19	.00	.00
9	8 0-30, 8 0-30	.00	.00
10	8 over 30, 8 over 30	.00	.00
11	10 0-30, 10 0-30	.00	.00
12	10 over 30, 10 over 30	.00	.00
13	12 0-30, 12 0-30	.237027	.119918
14	12 over 30, 12 over 30	.502538	.119956

Table 8.14: Conditional failure probabilities - days with exactly 2 breaks: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
15	6 0-19, 6 0-19	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.00	.00
18	10 0-30, 10 0-30, 10 0-30	.00	.00
19	10 over 30, 10 over 30, 10 over 30	.056604	.084501
20	12 0-30, 12 0-30, 12 0-30	.00	.00
21	12 over 30, 12 over 30, 12 over 30	.634573	.119934

Table 8.15: Conditional failure probabilities - days with exactly 3 breaks: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
no break		.00	.00	.00
1	6 0-19	.oo	.00	.00
2	8 0-30	.00	.00	.00
3	8 over 30	.00	.00	.00
4	10 0-30	.00	.00	.00
5	10 over 30	.00	.00	.00
6	12 0-30	.000248	.000765	.000474
7	12 over 30	.00	.00	.00

Table 8.16: System failure probabilities - days with exactly 0 or 1 break: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
8	8 0-30, 8 0-30	.00	.00	.00
9	8 0-30, 8 0-30	.00	.00	.00
10	8 over 30, 8 over 30	.00	.00	.00
11	10 0-30, 10 0-30	.00	.00	.00
12	10 over 30, 10 over 30	.00	.00	.00
13	12 0-30, 12 0-30	.000426	.001097	.000724
14	12 over 30, 12 over 30	.000326	.000779	.000535

Table 8.17: System failure probabilities - days with exactly 2 breaks: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
15	6 0-19, 6 0-19, 6 0-19	.00	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.00	.00	.00
18	10 0-30, 10 0-30, 10 0-30	.00	.00	.00
19	10 over 30, 10 over 30, 10 over 30	.0000212	.00019	.0000939
20	12 0-30, 12 0-30, 12 0-30	.00	.00	.00
21	12 over 30, 12 over 30, 12 over 3 0	.000279	.000666	.000459

Table 8.18: System failure probabilities - days with exactly 3 breaks: Lee et al. network

8.2 Rating the Solutions

In the previous sections of this chapter, Bonferroni confidence intervals for the probability of system failure solution were derived for each failure case in each network. The standard deviation and confidence limits at the .99 significance level are estimated using Eq. 5.9 for the cumulative probability of system failure for each network. The standard deviation of the cumulative probability of system failure is calculated from the variance for each network. These confidence intervals and the standard deviations are summarized in Table 8.25. In addition, least square, point estimates for the probability of system failure are also presented.

Table 8.25 shows that all networks are quite reliable with respect to the connectivity criterion. For all networks the confidence intervals do not overlap zero, thereby implying a non-zero failure probability. The probability of failure for all networks is, however, quite low and could be as low as 0.001051 at the .99 level of significance for the Morgan and Goulter network. On an intuitive basis, all solutions appear to be satisfactory, with respect to the connectivity criterion.

The Morgan and Goulter network, however, has a lower expected probability of failure than the other networks. The hypothesis that the Morgan and Goulter network is more reliable than the other networks is tested by comparing differences between the expected value of system failure for the Morgan and Goulter network with the expected value for each of the other networks. Since simultaneous comparisons are made Bonferroni techniques are used. The test statistic for these tests is:

$$z = \frac{TF_{other} - TF_{Morgan}}{\sqrt{S^2_{other} + S^2_{Morgan}}} \tag{8.1}$$

where

- TF_{other} = expected cumulative failure probability over all failure cases for one

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
no break		.00	.00
1	6 0-19	.00	.00
2	8 0-30	.00	.00
3	8 over 30	.00	.00
4	10 0-30	.00	.00
5	10 over 30	.00	.00
6	12 0-30	.00	.00
7	12 over 30	.316916	.111673

Table 8.19: Conditional failure probabilities - days with exactly 0 or 1 break: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
8	6 0-19, 6 0-19	.00	.00
9	8 0-30, 8 0-30	.00	.00
10	8 over 30, 8 over 30	.018405	.079271
11	10 0-30, 10 0-30	.00	.00
12	10 over 30, 10 over 30	.00	.00
13	12 0-30, 12 0-30	.00	.00
14	12 over 30, 12 over 30	.558568	.119954

Table 8.20: Conditional failure probabilities - days with exactly 2 breaks: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
15	6 0-19, 6 0-19, 6 0-19	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.080392	.088315
18	10 0-30, 10 0-30, 10 0-30	.00	.00
19	10 over 30, 10 over 30, 10 over 30	.00	.00
20	12 0-30, 12 0-30, 12 0-30	.00	.00
21	12 over 30, 12 over 30, 12 over 30	.697085	.119999

Table 8.21: Conditional failure probabilities - days with exactly 3 breaks: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
no break		.00	.00	.00
1	6 0-19	.00	.00	.00
2	8 0-30	.00	.00	.00
3	8 over 30	.00	.00	.00
4	10 0-30	.00	.00	.00
5	10 over 30	.00	.00	.00
6	12 0-30	.00	.00	.00
7	12 over 30	.000126	.000348	.000225

Table 8.22: System failure probabilities - days with exactly 0 or 1 break: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
8	6 0-19, 6 0-19	.00	.00	.00
9	8 0-30, 8 0-30	.00	.00	.00
10	8 over 30, 8 over 30	.00	.000202	.00000601
11	10 0-30, 10 0-30	.00	.00	.00
12	10 over 30, 10 over 30	.00	.00	.00
13	12 0-30, 12 0-30	.00	.00	.00
14	12 over 30, 12 over 30	.000521	.00112	.000845

Table 8.23: System failure probabilities - days with exactly 2 breaks: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
15	6 0-19, 6 0-19, 6 0-19	.00	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.000061	.000238	.000143
18	10 0-30, 10 0-30, 10 0-30	.00	.00	.00
19	10 over 30, 10 over 30, 10 over 30	.00	.00	.00
20	12 0-30, 12 0-30, 12 0-30	.00	.00	.00
21	12 over 30, 12 over 30, 12 over 3 0	.000442	.001037	.000721

Table 8.24: System failure probabilities - days with exactly 3 breaks: Morgan and Goulter network

Network	Bounds		Expected Value	Standard Deviation S	Number of Observations
	Lower Limit	Upper Limit			
Non-Optimal	-.00167	.007632	.002981	.001806	19
Gessler	.000998	.001786	.001392	.000153	22
Lee et al.	.001761	0.002811	0.002286	0.000204	22
Morgan and Goulter	.000774	.001771	.001273	0.000193	22

Table 8.25: Confidence intervals for cumulative probability of system failure

Comparison Network	Difference Between Expected Failure Probabilities	Combined Standard Deviation	Test Statistic z
Non-Optimal	.00171	.001816	.9415
Gessler	.00012	.000246	.4872
Lee et al.	.00101	.000281	3.5965

Table 8.26: Comparison of Morgan and Goulter network to other networks

of the other three networks

- TF_{Morgan} = expected cumulative failure probability over all failure cases for the Morgan and Goulter network
- S_{other} = standard deviation of the cumulative probability of failure for one of the other three networks
- S_{Morgan} = standard deviation of the cumulative probability of failure for the Morgan and Goulter network

The results for the three comparisons are shown in Table 8.26. The critical value for all tests is 2.807. Examination of Table 8.26 shows that the test statistic is greater than the critical value for only Lee et al. network. The test statistic for the the non-optimal and Gessler networks are less than the critical value. Thus, in spite of the small absolute differences in network reliability shown in Table 8.25, when the connectivity criterion is used the conclusion can be drawn that the Morgan and Goulter network is more reliable than Lee et al. network. The reliability levels of both the non-optimal and the Gessler networks in terms of the connectivity criterion is indistinguishable from that of the Morgan and Goulter network.

It should be noted, however, that the evaluation of a network's connectivity is a necessary step in determining the network's hydraulic performance. In order to consider the hydraulic performance aspects, the computer runs used to determine network reliability with respect to connectivity were combined with the computer runs to evaluate network performance, based on hydraulic performance. The reliability of these networks, based on hydraulic performance, is discussed in the next chapter.

Chapter 9

Estimation of Reliability Performance of Networks using Hydraulic Criteria

In this chapter, the performance of the test networks with respect to hydraulic performance is discussed. Satisfactory hydraulic performance of the network is defined as the provision of specified flows to each node at, or above, specified pressure levels. A system failure is deemed to have occurred when the system fails to meet the specified flows and pressures at any node. This hydraulic failure criterion is far more stringent than the connectivity criterion used in the previous chapter.

The required nodal flows are specified as 1.3 times the average daily demand as specified in Table 7.2. The required pressure levels are also specified in Table 7.2. These flows correspond to the period of the highest average daily demand level. This demand level is chosen because it is the highest average daily demand level which can be modeled based of the information specified in Walski et al. (1987) and because it is the highest level of demand that can be expected to occur on a daily basis. Thus, assuming that 24 hours is required to fix pipe breaks, this demand level is likely to occur during any pipe breakage event. It should be noted that higher demand levels which occur less frequently require the simultaneous occurrence of a low probability pipe failure event and a low probability high demand event. The combined probability of two such events tends to be quite small reflecting the general rarity of the event.

In Table 7.4, the duration of this demand is assumed to be three hours. Since this demand level lasts for three hours, the possibility exists that overhead storage tanks could be emptied. As discussed in Chapter 8, flow constraints are imposed on the flow, from each overhead tank, to prevent the tanks from being emptied.

The basic method, for estimating network reliability, is the same as the method used in Chapter 8. The use of hydraulic performance criteria, rather than connectivity, makes the estimation of the conditional probability term, $P(F \mid RE_i)$, in Eq. 5.6 more complex and more computationally intensive.

9.1 Changes to the Monte Carlo Simulation Algorithm

Monte Carlo simulation is used to estimate the probability that the network fails when links are removed. For hydraulic performance criteria, the simulation program works as follows:

1. A failure case which specifies the pipe categories in which pipe breaks are to occur is selected.

2. A pipe is selected randomly for each pipe category in the failure case. The pipe selections are weighted by the length of the pipe, with respect to the total length of pipe, in the pipe category. If a pipe category contains 10 km of pipe

and a particular pipe has a length of 1 km then that pipe would be chosen by the program with a probability of .10.

3. The selected pipes are removed from the network.

4. The network is checked for disconnected nodes. It is assumed that if a demand node is disconnected from the source node then the demand node receives no flow and, thus, a system failure has occurred. In this case no hydraulic analysis of the network is performed.

5. If the network is connected, it is solved hydraulically (using the linear theory technique).

6. Flow and pressure levels at each node are checked against required pressures and flows.

7. The proportion of removal sets which cause disconnection and/or network failure with respect to pressure and flow conditions is determined by repeated trials. The error in the estimate is then determined, using Eq. 4.22.

8. Trials are repeated until the error is reduced to a preset level.

In general, this algorithm is quite similar to the one presented for the connectivity failure criterion in Chapter 8. A number of technical difficulties must, however, be faced in implementing the algorithm using hydraulic performance criteria. The first difficulty is that solving a hydraulic network is a non-linear problem. This non-linearity can add greatly to the required computational effort.

The second difficulty arises because the solution of a hydraulic network requires that a set of simultaneous equations be set up and solved. These equations depend on the physical layout of the network. Specifically, equations must be formulated to represent the hydraulic head loss around loops in the network. For consistency, these equations require that a rigid, and somewhat arbitrary, sign convention be followed. Pipes in a loop, in which flow is assumed to travel in a clockwise direction around the loop, are assigned a positive sign. The practice described in the literature (Epp and Fowler, 1970; Wood and Charles, 1972; Jeppson and Tavallee, 1975; and Jeppson and Davis, 1976) is to specify the loop equations as input to the network solver.

This practice is sensible because, using a network layout diagram, it is relatively easy to identify loops and to identify the clockwise direction, for each loop. Unfortunately, when links are removed from a network, at random, during the Monte Carlo simulation process, the network configuration is also changed. For this reason, it is necessary to reconfigure the network each time a set of links is removed. Fowler and Epp (1970) present an algorithm to identify loops in a network. This algorithm is not sufficiently general to identify all necessary loops in the network used for this study. The algorithm therefore was extended, and computer code developed, to implement the required sign convention. This computer code makes it possible to dynamically set up a system of equations, for each randomly generated network configuration.

Three primary methods of solving the equations are presented in the literature. The Hardy-Cross method, the Newton-Raphson method (Fowler and Epp, 1970), and the linear theory method (Wood and Charles, 1972). Both the Hardy-Cross method and the Newton-Raphson method require good estimates of the flows in each pipe in the network to ensure rapid convergence. Wood and Charles (1972) point out that very bad initial estimates of these flows can lead to very slow convergence or even non-convergence. The requirement of good initial estimates of pipe flows is a serious drawback for use in Monte Carlo simulation as it is difficult to develop algorithms which guarantee good initial estimates of pipe flows.

The third method of Wood and Charles (1972) is based on linear theory and does not require good initial estimates of link flows. The only disadvantage, of this method, is that it requires more computer memory than either the Hardy-Cross or the Newton-Raphson methods. However, this need for greater computer memory is not a serious issue as the program was implemented on a Sun workstation in a UNIX environment and used far less memory than the maximum available on that system. Furthermore, using this method, solutions were obtained for all network configurations generated by the Monte Carlo simulation is less than 50 iterations.

The probabilities of network failure based on hydraulic sufficiency and the confidence limits on the probability of system failure were then calculated for each of the four networks discussed in Chapter 7. The results of these calculations of probability of failure and the associated confidence limits are discussed in the following section.

9.2 The Non-Optimal Network

The Bonferroni confidence limits for the probability of occurrence for each failure case in the covering set for this network are unchanged from those presented in Tables 7.22, 7.23, and 7.24. The estimated probability of hydraulic network failure, given that a removal event has occurred, is shown for each failure case in Tables 9.1, 9.2, and 9.3. The estimated error associated with the probability estimates is also shown in these tables. Upper bounds for the probability of network failure, based on hydraulic criteria, for each case are calculated using Eq. 5.9. Lower bounds for the probability of failure, for each case are calculated using Eq. 5.10. Upper bounds for estimated error with respect to these estimates are calculated using Eq. 5.12. These results are shown in Tables 9.4, 9.5, and 9.6.

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
no break		.00	.00
1	8 0-30	.256729	.049996
2	8 over 30	.00	.00
3	10 0-30	.00	.00
4	10 over 30	.00	.00
5	12 0-30	.171658	.049998
6	12 over 30	.89576	.094214

Table 9.1: Conditional failure probabilities - days with exactly 0 or 1 break: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
7	8 0-30, 8 0-30	.303137	.049992
8	8 over 30, 8 over 30	.00	.00
9	10 0-30, 10 0-30	.113199	.049983
10	10 over 30, 10 over 30	.00	.00
11	12 0-30, 12 0-30	.366181	.049997
12	12 over 30, 12 over 30	.898104	.07638

Table 9.2: Conditional failure probabilities - days with exactly 2 breaks: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
13	8 0-30, 8 0-30, 8 0-30	.388283	.049995
14	8 over 30, 8 over 30, 8 over 30	.00	.00
15	10 0-30, 10 0-30, 10 0-30	.188164	.049988
16	10 over 30, 10 over 30, 10 over 30	.00	.00
17	12 0-30, 12 0-30, 12 0-30	.564185	.049997
18	12 over 30, 12 over 30, 12 over 30	.944471	.055529

Table 9.3: Conditional failure probabilities - days with exactly 3 breaks: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
no break		.00	.00	.00
1	8 0-30	.000394	.000497	.000445
2	8 over 30	.00	.00	.00
3	10 0-30	.00	.00	.00
4	10 over 30	.00	.00	.00
5	12 0-30	.00039	.00080	.000581
6	12 over 30	-.0000069	.0000191	.00000176

Table 9.4: System failure probabilities - days with exactly 0 or 1 break: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
7	8 0-30, 8 0-30	.000212	.000296	.000252
8	8 over 30, 8 over 30	.00	.00	.00
9	10 0-30, 10 0-30	.000148	.000281	.000211
10	10 over 30, 10 over 30	.00	.00	.00
11	12 0-30, 12 0-30	.00094	.001588	.001252
12	12 over 30, 12 over 30	.001343	.002702	.002006

Table 9.5: System failure probabilities - days with exactly 2 breaks: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
13	8 0-30, 8 0-30, 8 0-30	.000124	.000145	.000136
14	8 over 30, 8 over 30, 8 over 30	.00	.00	.00
15	10 0-30, 10 0-30, 10 0-30	.000144	.000153	.000153
16	10 over 30, 10 over 30, 10 over 30	.00	.00	.00
17	12 0-30, 12 0-30, 12 0-30	.000879	.001534	.001199
18	12 over 30, 12 over 30, 12 over 3 0	.000955	.001978	.001460

Table 9.6: System failure probabilities - days with exactly 3 breaks: non-optimal network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
no break		.00	.00
1	6 0-19	.00	.00
2	8 0-30	.00	.00
3	8 over 30	.00	.00
4	10 0-30	.00	.00
5	10 over 30	.480769	.119947
6	12 0-30	.00	.00
7	12 over 30	.649061	0.119936

Table 9.7: Conditional failure probabilities - days with exactly 0 or 1 break: Gessler network

9.3 The Gessler Network

The analysis for the Gessler solution is similar to that used for the non-optimal network. The estimated probability of network failure, given that a removal event has occurred, is shown for each failure case in Tables 9.7, 9.8, and 9.9 with the bounds for the probability of system failure being shown in Tables 9.10, 9.11, and 9.12.

9.4 The Lee et al. Network

The analysis for the Lee et al. solution is, again, similar to that used for the non-optimal network. The estimated probability of network failure, given that a removal event has occurred, is shown for each failure case in Tables 9.13, 9.14, and 9.15 with the bounds for the probability of system failure being shown in Tables 9.16, 9.17, and

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
8	6 0-19, 6 0-19	.00	.00
9	8 0-30, 8 0-30	.00	.00
10	8 over 30, 8 over 30	.00	.00
11	10 0-30, 10 0-30	.00	.00
12	10 over 30, 10 over 30	.47923	.119993
13	12 0-30, 12 0-30	.00	.00
14	12 over 30, 12 over 30	.737101	.113177

Table 9.8: Conditional failure probabilities - days with exactly 2 breaks: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
15	6 0-19, 6 0-19, 6 0-19	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.00	.00
18	10 0-30, 10 0-30, 10 0-30	.00	.00
19	10 over 30, 10 over 30, 10 over 30	.473176	.119963
20	12 0-30, 12 0-30, 12 0-30	.00	.00
21	12 over 30, 12 over 30, 12 over 30	.817495	.093108

Table 9.9: Conditional failure probabilities - days with exactly 3 breaks: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
no break		.00	.00	.00
1	6 0-19	.00	.00	.00
2	8 0-30	.00	.00	.00
3	8 over 30	.00	.00	.00
4	10 0-30	.00	.00	.00
5	10 over 30	.001901	.003281	.002546
6	12 0-30	.00	.00	.00
7	12 over 30	.000331	.000771	.000535

Table 9.10: System failure probabilities - days with exactly 0 or 1 break: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
8	6 0-19, 6 0-19	.00	.00	.00
9	8 0-30, 8 0-30	.00	.00	.00
10	8 over 30, 8 over 30	.00	.00	.00
11	10 0-30, 10 0-30	.00	.00	.00
12	10 over 30, 10 over 30	.001484	.002602	.002004
13	12 0-30, 12 0-30	.00	.00	.00
14	12 over 30, 12 over 30	.000827	.001826	.001298

Table 9.11: System failure probabilities - days with exactly 2 breaks: Gessler network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability Lower Limit	Upper Limit	Expected Value
15	6 0-19, 6 0-19, 6 0-19	.00	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.00	.00	.00
18	10 0-30, 10 0-30, 10 0-30	.00	.00	.00
19	10 over 30, 10 over 30, 10 over 30	.000731	.00132	.001004
20	12 0-30, 12 0-30, 12 0-30	.00	.00	.00
21	12 over 30, 12 over 30, 12 over 3 0	.000621	.001379	.000984

Table 9.12: System failure probabilities - days with exactly 3 breaks: Gessler network 9.18.

9.5 The Morgan and Goulter Network

The analysis for the Morgan and Goulter solution is, again, similar to that used for the non-optimal network. The estimated probability of network failure, given that a removal event has occurred, is shown for each failure case in Tables 9.19, 9.20, and 9.21 with the bounds for the probability of system failure being shown in Tables 9.22, 9.23, and 9.24.

Confidence intervals for these four solutions are presented in Table 9.25.

9.6 Discussion of Results

The expected probability of network failure shown in Table 9.25 is higher for all networks than the expected probability of network failure shown in Table 8.25. The

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
no break		.00	.00
1	6 0-19	.00	.00
2	8 0-30	.00	.00
3	8 over 30	.00	.00
4	10 0-30	.00	.00
5	10 over 30	.00	.00
6	12 0-30	.784962	.119962
7	12 over 30	.00	.00

Table 9.13: Conditional failure probabilities - days with exactly 0 or 1 break: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
8	6 0-19, 6 0-19	.00	.00
9	8 0-30, 8 0-30	.00	.00
10	8 over 30, 8 over 30	.00	.00
11	10 0-30, 10 0-30	.00	.00
12	10 over 30, 10 over 30	.00	.00
13	12 0-30, 12 0-30	.908836	.081168
14	12 over 30, 12 over 30	.782741	.098936

Table 9.14: Conditional failure probabilities - days with exactly 2 breaks: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
15	6 0-19, 6 0-19, 6 0-19	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.00	.00
18	10 0-30, 10 0-30, 10 0-30	.00	.00
19	10 over 30, 10 over 30, 10 over 30	.122642	.11995
20	12 0-30, 12 0-30, 12 0-30	.00	.00
21	12 over 30, 12 over 30, 12 over 30	.869803	.083813

Table 9.15: Conditional failure probabilities - days with exactly 3 breaks: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
no break		.00	.00	.00
1	6 0-19	.00	.00	.00
2	8 0-30	.00	.00	.00
3	8 over 30	.00	.00	.00
4	10 0-30	.00	.00	.00
5	10 over 30	.00	.00	.00
6	12 0-30	.00172	.003101	.002379
7	12 over 30	.00	.00	.00

Table 9.16: System failure probabilities - days with exactly 0 or 1 break: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
8	6 0-19, 6 0-19	.00	.00	.00
9	8 0-30, 8 0-30	.00	.00	.00
10	8 over 30, 8 over 30	.00	.00	.00
11	10 0-30, 10 0-30	.00	.00	.00
12	10 over 30, 10 over 30	.00	.00	.00
13	12 0-30, 12 0-30	.002077	.003522	.002776
14	12 over 30, 12 over 30	.000539	.001154	.000833

Table 9.17: System failure probabilities - days with exactly 2 breaks: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
15	6 0-19, 6 0-19, 6 0-19	.00	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.00	.00	.00
18	10 0-30, 10 0-30, 10 0-30	.00	.00	.00
19	10 over 30, 10 over 30, 10 over 30	.0000884	.000351	.000204
20	12 0-30, 12 0-30, 12 0-30	.00	.00	.00
21	12 over 30, 12 over 30, 12 over 3 0	.000402	.000875	.000630

Table 9.18: System failure probabilities - days with exactly 3 breaks: Lee et al. network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
no break		.00	.00
1	6 0-19	.00	.00
2	8 0-30	.00	.00
3	8 over 30	.00	.00
4	10 0-30	.00	.00
5	10 over 30	.00	.00
6	12 0-30	.00	.00
7	12 over 30	.510707	0.11998

Table 9.19: Conditional failure probabilities - days with exactly 0 or 1 break: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
8	6 0-19, 0-19	.00	.00
9	8 0-30, 8 0-30	.00	.00
10	8 over 30, 8 over 30	.042945	.119565
11	10 0-30, 10 0-30	.00	.00
12	10 over 30, 10 over 30	.00	.00
13	12 0-30, 12 0-30	.00	.00
14	12 over 30, 12 over 30	.755965	.103758

Table 9.20: Conditional failure probabilities - days with exactly 2 breaks: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	Conditional Probability of System Failure Given Removal ($\hat{p}$)	Error Estimate for $\hat{p}$
15	6 0-19, 6 0-19, 6 0-19	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.162745	.119897
18	10 0-30, 10 0-30, 10 0-30	.00	.00
19	10 over 30, 10 over 30, 10 over 30	.00	.00
20	12 0-30, 12 0-30, 12 0-30	.00	.00
21	12 over 30, 12 over 30, 12 over 30	.885932	.083015

Table 9.21: Conditional failure probabilities - days with exactly 3 breaks: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
no break		.00	.00	.00
1	6 0-19	.00	.00	.00
2	8 0-30	.00	.00	.00
3	8 over 30	.00	.00	.00
4	10 0-30	.00	.00	.00
5	10 over 30	.00	.00	.00
6	12 0-30	.00	.00	.00
7	12 over 30	.000218	.000533	.000362

Table 9.22: System failure probabilities - days with exactly 0 or 1 break: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
8	6 0-19, 6 0-19	.00	.00	.00
9	8 0-30, 8 0-30	.00	.00	.00
10	8 over 30, 8 over 30	-.000046	.000359	.000140
11	10 0-30, 10 0-30	.00	.00	.00
12	10 over 30, 10 over 30	.00	.00	.00
13	12 0-30, 12 0-30	.00	.00	.00
14	12 over 30, 12 over 30	.000735	.001595	.001143

Table 9.23: System failure probabilities - days with exactly 2 breaks: Morgan and Goulter network

Failure Case	Failure Pattern Consists of 1 Pipe Removed From Each Pipe Category Shown	System Failure Probability		
		Lower Limit	Upper Limit	Expected Value
15	6 0-19, 6 0-19, 6 0-19	.00	.00	.00
16	8 0-30, 8 0-30, 8 0-30	.00	.00	.00
17	8 over 30, 8 over 30, 8 over 30	.00017	.000426	.000289
18	10 0-30, 10 0-30, 10 0-30	.00	.00	.00
19	10 over 30, 10 over 30, 10 over 30	.00	.00	.00
20	12 0-30, 12 0-30, 12 0-30	.00	.00	.00
21	12 over 30, 12 over 30, 12 over 3 0	.000585	.001271	.000916

Table 9.24: System failure probabilities - days with exactly 3 breaks: Morgan and Goulter network

Network	Bounds		Expected Value	Standard Deviation	Number of Observations
	Lower Limit	Upper Limit			
Non-Optimal	-.018900	0.017282	0.007695	.00372	19
Gessler	0.003402	0.005697	0.00455	.000445	22
Lee et al.	0.005740	0.007901	0.00682	.000419	22
Morgan and Goulter	.001315	0.002556	0.001935	.000241	22

Table 9.25: Confidence intervals for cumulative probability of system failure

higher failure probabilities occur because Table 9.25 is based on hydraulic failure criteria rather than the connectivity failure criterion. It is possible for a node to be connected while receiving inadequate flow at inadequate pressure levels. On the other hand if a node is disconnected from the network it receives no flow. Thus the use of hydraulic failure criteria is more conservative than the use of the connectivity criterion. It should be noted that because the probabilities given in Table 9.25 are daily failure probabilities the expected number of days in a year in which on inadequate service levels delivered varies from .7 days for the Morgan and Goulter network to 2.8 days for the non-optimal network. However, by the non-optimal nature of the manner in which the 'non-optimal' network was designed the non-optimal network is also more expensive to implement than the Morgan and Goulter solution.

On an intuitive basis, the results in Table 9.25 suggest that, based on hydraulic criteria, the Morgan and Goulter network is again the most reliable solution. The hypothesis that the Morgan and Goulter network is more reliable than the other networks was tested in a similar fashion to the method used in Chapter 8. Table 9.26 contains the results of the hypothesis testing used to compare the Morgan and Goulter network to the other networks. The critical value for all tests is 2.807.

Examination of Table 9.26 shows that the test statistic is greater than the critical

Comparison Network	Difference Between Expected Failure Probabilities	Combined Standard Deviation	Test Statistic z
Non-Optimal	.00576	.00373	1.545
Gessler	.00262	.000506	5.178
Lee et al.	.00489	.000483	10.117

Table 9.26: Comparison of Morgan and Goulter network to other networks

value for the Gessler and Lee et al. networks. The test statistic for the non-optimal network is not significant. It is interesting to note that while the test statistic for the Gessler network is less than the critical value when the connectivity criterion is used, the corresponding test statistic for hydraulic failure criteria is statistically significant. Thus, the conclusion can be drawn based on hydraulic failure criteria that the Morgan and Goulter network is more reliable than the Gessler and Lee et al. networks, but is not distinguishable from the non-optimal network. These results show that it is possible to rank the networks according to reliability.

9.7 Computational Requirements

Since, as noted earlier, the exact calculation of reliablity in networks is NP hard it is useful to review computational requirements of this approximate method. The computer runs, to estimate the reliability of the four solutions, were all made on Sun SPARC-2 work stations, running under UNIX. These work stations are capable of running several jobs simultaneously. During the runs, the program, which estimated reliability, typically used 30 - 50% of the machine capacity. The run times, for the four networks, are presented in Table 9.27. Two runs are shown for the non-optimal network. The first run was made using a small simulation error for the conditional

Solution	Run Time (min)	Simulation error
Non-optimal(.005)	3848	.005
Non-optimal(.12)	642	.12
Gessler	287	.12
Lee et al.	414	.12
Morgan and Goulter	61	.12

Table 9.27: Elapsed running times and associated errors for Monte Carlo Simulation Runs

probality of network failure given the removal of a set of links (.005). The second run was made using a larger allowable simulation error (.12). The larger simulation error was still found to produce acceptable results. No trials were made for the other networks for the smaller error because of the excessive computational effort required (3848 minutes is 2.67 days). It should be noted that, the computational effort required to determine network reliability is determined by the requirements of the Monte Carlo simulation step. This issue is an important contribution to the computational requirements of the methodology. The estimation of component reliabilities is performed only once for a given network, while the simulation step must be performed for each possible network configuration.

It can be seen from Eq. 5.6 that several factors affect the computational requirements for estimating network reliability. The most important of these factors are the level of required precision, the size of the conditional reliability terms, the complexity of the network, and the size of the network.

The importance of the required precision level is illustrated by the excessive running times, used for the non-optimal network when the required precisionwas 0.005. The running time for the 'non-optimal' network was much lower when a required precision of 0.12 was used.

The size of the conditional reliability terms affects the running times because the simulation step estimates a proportion using a binomial process. The error in the estimate of a proportion is dependent on the variance of the proportion as given by Eq. 4.21. The numerator of the variance is, in turn, dependent on the value of the proportion. The variance has a maximum at $\hat{p} = 0.50$ and minima at $\hat{p} = 0$ and 1.0. The relatively low computational requirement for the Morgan and Goulter solution is, in large part, due to the low values for the conditional failure probabilities for all cases tested.

The complexity of the network depends on the number of different pipe categories included in a particular solution. The complexity of the network affects the $Z_{\beta/2}$ value, in Eq. 4.22, because a more complex network requires the simultaneous estimation of more parameters. Recall that Bonferroni techniques require that the required significance level increases as the number of parameters that must estimated simultaneously increases. The increased significance level requires an increased value of $Z_{\beta/2}$. As the complexity of the solution networks did not vary greatly, this factor had only a minimal effect on the relative computational requirements for the networks considered in this thesis.

Finally, the size of the network affects both the number of failure cases, that must be considered, and the time necessary to solve the network at each iteration.

Chapter 10

Discussion of Results

10.1 Evaluation of Solution Performance

In this thesis a method was presented to estimate water supply distribution network reliability. This method is based on the separate estimation of pipe failure probabilities and system failure probabilities given that combinations of pipes have failed. The method is different from methods presented in the literature in that pipe failure probabilities are not assumed, but are estimated using statistical techniques. The conditional probabilities of system failure given that various combinations of pipes have been removed is also estimated statistically using Monte Carlo simulation techniques. Although the method only estimates reliability, acceptable results were produced for all test cases. The error in the system reliability estimates is less than .01 for all test cases. It was possible to distinguish between the levels of reliability for most of the test networks using hypothesis testing at high levels of statistical significance.

Primary considerations in the application of this methodology are the computational effort required and the error in the reliability estimates produced. The factors which influence the computational effort and the error in the reliability estimates are discussed in the following sections of this chapter.

10.2 Failure Criteria

Two criteria for network failure were examined in this thesis: connectivity and hydraulic performance. All of the networks appeared to be quite reliable with respect to the connectivity failure criterion. In terms of this failure criterion the Morgan and Goulter network was found to be significantly more reliable than the Lee et al. network. The non-optimized network and the Gessler network were not found to be significantly different from the Morgan and Goulter network using the connectivity criteria.

A primary advantage of the connectivity criterion, over hydraulic performance criteria, is the smaller computational burden associated with the connectivity criterion. Two factors make the estimation of network reliability using the connectivity criterion less computationally intensive than the estimation of reliability using the hydraulic criteria. The first is that the evaluation of network connectivity requires less effort than solving the network hydraulically. The second factor is the high degree of reliability of these networks with respect to connectivity. As discussed previously, the required number of iterations, to achieve a given degree of precision, is dependent on the value of $\hat{p}$, with minima at 0 and 1.0. Since probability of node disconnection is close to 0 for all cases very few iterations are required. However, the main weakness of the use of connectivity as a failure criterion is that it is non-conservative. The existence of an unbroken path from the source pumps to a user's water intake does not ensure delivery of water to the user. As a result, despite its relative simplicity, connectivity has serious drawbacks as a measure of network reliability.

The second failure criteria, namely, the ability of the network to provide adequate pressure and flow levels at all nodes during peak demand periods is more realistic. As a result, the hydraulic performance criterion, used in this thesis, is a more valid measure of system performance. The reliability of the solutions, with respect to the hydraulic criteria, varies widely. In particular, the non-optimal and the Lee et al. networks could be expected to provide inadequate levels of service .77% and .68% (2.8 and 2.5 days) of the days in the year, respectively. Further investigation to determine the nodes that are inadequately served and possible remedial action seems reasonable for these networks. The results in Table 9.26 show that the Morgan and Goulter solution is more reliable than the Lee et al. and the Gessler networks. The t-value for the Non-optimal network is not statistically significant. As a result the reliability of the Non-optimal network cannot be distinguished from that of the Morgan and Goulter Network. However, the greater cost inherent in its design shows that a 'non-optimal' replacement of pipe in large areas of a network at great cost is not necessarily a good method to achieve required reliability levels.

It should be noted that other operational conditions may be critical. These conditions include fire flows, instantaneous peak demands, and the filling of overhead storage tanks during off peak hours.

10.3 Sources of Error

From Eq. 5.13, two sources of error, in the estimate of system reliability, can be determined. The first source is the error that occurs due to lack of knowledge or variability of the failure behaviour of network components. The second source is the error in the simulation process. It should be noted that the error due to simulation is much smaller than the total error for all cases. In all cases, most of the error is due to uncertainty with respect to component failure behaviour. In view of this result, and the rather large computational cost associated with decreasing the level of simulation error, there is little justification for the extra effort required to reduce the simulation error.

As noted above the error level in the estimate of system reliability is, to a large extent, determined by factors such as the variability of the reliability of system components, and the probability of system failure, given that some system components have failed. For a given network configuration, these sources of error are, to a large extent, uncontrollable.

From a research perspective, the development of a simpler model, to predict failure behaviour, could greatly reduce the error associated with network failure probability estimates. The model developed in this thesis depends on pipe material, pipe age, and pipe diameter. The results suggest that there is a relationship between pipe diameter and failure probability. However, markedly different failure rates were found for pipes with identical pipe material and diameter but different pipe ages. The existence of a sudden large change in failure rate for pipes of differing pipe ages is not caused only by a continuous aging process. The change in failure rate is in part due to changes in pipe construction. In particular, component failure probabilities were found to vary for for cast iron pipes of the same diameter because the wall thickness of cast iron pipe has been decreased in the newer pipe. If a model using pipe wall thickness, rather than pipe diameter, as an independent variable could be developed it might be more accurate and less complex than the model developed in this work. The reduction in complexity would come from the use of a single variable, namely, pipe wall thickness, rather multiple variables, to represent every pipe diameter. The reduction in the number of variables would, in turn, reduce Bonferroni's g. The smaller Bonferroni's g would make the size of the confidence intervals smaller, and thus reduce the error in the estimate of network failure probability.

The method used to estimate component reliabilities is quite complex and involves a large number of assumptions. The method is, however, accurate enough to demonstrate that system reliability can be estimated with reasonable levels of accuracy if the probability distribution for component reliabilities is known. In effect the method for estimating system reliability for water distrbution systems presented in this thesis has reduced the system reliability problem to the less complex problem of estimation of component reliabilities.

10.4 Effects of System Size and Complexity

The computational effort associated with the technique increases with system size for two reasons. The first reason is that a bigger system has more pipes and takes longer to solve. As a result, each iteration takes longer for a larger system than for a smaller system. It should be noted that the network used in this study contains approximately 81 km of pipe, and so is a fairly large network. Walski (1987) states that this network is designed to include most of the characteristics of a medium sized city, without the detailed distribution grid. The processing times for all networks are reasonable. Larger networks could be handled as long as the number of failure cases to be processed does not increase substantially.

The second reason that computational effort increases with system size is that more simultaneous failures are likely to occur in a larger network. This effect, in turn, increases the number of failure cases to be considered. It should be noted that the number of simultaneous failures that is likely to occur in 24 hours does not increase very rapidly. Jacobs and Goulter (1991) show that for the entire City of Winnipeg, which contains 2259 km of pipe, even with the significant clustering of breaks reported in that city by Goulter and Kazemi (1988)(1989) there is less than a 0.01 probability of more than 20 breaks occurring in a single day. This result suggests that the number of simultaneous breaks that must be considered increases quite slowly, with system size.

As a result of the above considerations, it should be reasonable to estimate network reliability for systems two or three times bigger than the ones used in this study, if hydraulic performance is used as the network failure criterion. If connectivity is used as the network failure condition, then the estimation of failure probabilities for networks that are ten times the size of the one used in this study are possible.

It should be noted that the network described in Walski et al. (1987) has one feature that is quite unrealistic. The network uses only cast iron pipe. The City of Winnipeg system contains cast iron, asbestos cement, PVC, and ductile steel pipe, in relatively large quantities. In this sense, distribution networks found in practice are more complex than the one used in this study. The effect of a more complex network is that more parameters must be estimated. The increased number of parameters to be estimated, increases the uncertainty of the estimate. The method used, in this study, is based on the statistical estimation of a large number of parameters. An increase in network complexity could decrease the level of statistical accuracy to unacceptable levels. Future research should investigate methods of limiting the number of parameters used to describe failure behaviour. In addition, future research concerning more efficient statistical methods of estimating error would be valuable.

10.5 Mean Time to Repair

The assumption was made that the time to repair pipe breaks in the City of Winnipeg is less than 24 hours. This assumption is reasonable for current operational conditions in the City of Winnipeg. The mean time to repair a pipe break directly affects system reliability because it determines the duration that a pipe remains out of service. In an ideal network, in which all breaks are independent, a doubling of the mean time to repair a break would also double the expected number of breaks that occur before the broken pipe is repaired. In a network in which pipe breaks tend to cluster in time and space the effects of doubling the mean time to repair would be even larger. The effects of increasing mean time to repair have not been explored in this thesis.

Further research into this question is necessary.

Chapter 11

Summary and Conclusions

The calculation of network reliability for water supply distribution systems is difficult. This makes it hard to evaluate the relative merits of proposed designs. It has been shown that the exact calculation of network reliability is so-called NP-hard. This means that the time required to perform exact reliability calculations for large networks is measured in years. For this reason the methods currently used in the design of reliable water supply networks are based on heuristic, surrogate measures of reliability. They have the advantage that they require far less computational effort than the exact calculation of network reliability, but the disadvantage that solutions based on each surrogate measure differ and that no method exists to compare the different results.

A method has been presented which uses sampling by Monte Carlo techniques to estimate network reliability. This method decomposes network reliability into the probability of component failure and the conditional probability of system failure given that a set of components has failed. The probability of component failure is estimated based on failure history records. The conditional probability of system failure given that components have failed is estimated using Monte Carlo simulation.

This method was applied to four existing solutions to the problem of designing a least cost, reliable pipe network for a particular pipe network. Three of these solutions were determined so as to meet surrogate measures of reliability. The fourth solution was found using trial and error and involved the replacement of almost all pipes over 50 years old. Network reliability was estimated to within 1% for all four solutions using two criteria to determine network failure: connectivity and hydraulic adequacy. Thus, it was possible to compare the reliability of the four networks. Highly statistically significant differences were found between the comparison of the Morgan-Goulter network and both the Gessler and Lee et al. networks using the hydraulic adequacy criterion. It was not possible to distinguish between the reliability of the Non-optimal network and the Morgan-Goulter Network. It should be noted that the greater cost inherent in the Non-optimal network makes it unattractive.

The computational effort required to estimate reliability by sampling was found to be large, but not excessive. It would not be feasible to include this method within an optimization method that requires the evaluation of network reliability a large number of times. The method can be used, however, to evaluate surrogate measures of reliability. In addition, the method could be used to check the reliability of any existing water distribution network design. As such this method can be a valuable design and research tool.

Further research should be aimed at reducing the required computational effort and improving the accuracy of error estimates. Promising areas of research are estimation of the conditional failure probabilities using Monte Carlo simulation and the failure process of the pipes in water supply distribution networks. In addition, the effect of the time required to repair pipe breaks requires further study.

Bibliography

[1] Agrawal, Avinash and Richard E. Barlow, (1984). 'A survey of network reliability and domination theory', *Operations Research*, 32(3), 478-492.

[2] Alperovits, E. and U. Shamir, (1977). 'Design of optimal water distribution systems', *Water Resources Research*, 13(6), 885-900.

[3] Arnold, G. E., (1960). 'Experience with main breaks in four large cities - Philadelphia', *Journal of the American Waterworks Association*, 52(8), 1041-1044.

[4] Andreou, S.A., Marks, D.H., and R.M. Clark, (1987). 'A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory', *Advances in Water Resources* 10(2), 2-10.

[5] Andreou, S.A., D.H. Marks, and R. M. Clark, (1987). 'A new methdology for modelling break failure patterns in deteriorating water distribution systems: Application', *Advances in Water Resources*, 10(2), 11-20.

[6] Awumah, Kofi, (1990). *Entropy Based Redundancy Measures in Water Distribution Network Design*, Ph.D. Thesis, University of Manitoba, Winnipeg, Manitoba, pp. 217.

[7] Awumah, Kofi, Goulter, Ian, and Suresh K. Bhatt, (1991). 'Entropy-based measures in water-distribution networks', *Journal of Hydraulic Engineering, ASCE*, 117(5), 595-614.

[8] Ball, Michael. O., (1980). 'The complexity of network reliability computations', *Networks*, 10(4), 153-165.

[9] Ball, Michael O. and George Nemhauser, (1979). 'Matroids and a reliability analysis problem', *Mathematics of Operations Research*, 4(2), 132-143.

[10] Ball, Michael O. and J. Scott Provan, (1983). 'Calculating bounds on reachability and connectedness in stochastic networks',*Networks*, 13(2), 253-278.

[11] Buzacott, J. A., (1980). 'A recursive algorithm for finding reliability measures related to the connection of nodes in a graph', *Networks*, 10(4) 311-327.

[12] Buzacott, J. A., (1983). 'A recursive algorithm for directed-graph reliability', *Networks*, 13(2), 241-246.

[13] Cenedese, A. and P. Mele, (1978). 'Optimal design of water distribution networks', *Journal of the Hydraulics Division, ASCE*, 104(HY2), 237-248.

[14] Clark and Goodrich, (1989)."Developing a data base on infrastructure needs",*Journal of the American Waterworks Association*, 81(7), 81-87.

[15] Coals A. and I. C. Goulter, (1985). 'Approaches to the consideration of reliability in water distribution networks', International Symposium on Urban Hydrology, Hydraulic Infrastructures, and Water Quality Control, University of Kentucky, Lexington, Kentucky, July 23-25, 287-295.

[16] Cross, H., (1936). 'Analysis of flow in networks of conduits or conductors', *Engineering Experiment Station Bulletin No. 286*, University of Illinois.

[17] Dixon, Wilfrid J. and Frank J Massey Jr., (1969), *Introduction to Statistical Analysis*, McGraw-Hill, New York, pp. 638, third edition.

[18] Elms, D. G., (1983). 'From a structure to a tree', *Civil Engineering Systems*, 1(1), 95-106.

[19] Endrenyi, J., (1978). *Reliability modeling in electric power systems.* John Wiley and Sons, New York, pp. 338.

[20] Epp, Robert and Alvin G. Fowler, (1970). 'Efficient code for steady-state flows in networks', *Journal of the Hydraulics Division ASCE*, 96(HY1), 43-56.

[21] Fishman, G. S., (1990). 'A comparison of four Monte Carlo Methods for estimating the probability of s-t connectedness', *IEEE Transactions on Reliability*, R-35(2), 145-155.

[22] Fujiwara, O., B. Jenchaimahakoon and N. C. P. Edirisinghe, (1970). 'A modified linear programming gradient method for optimal design of looped water water distribution networks', *Water Resources Research*, 23(6), 977-982.

[23] Goulter, I. C., (1987). 'Current and future use of systems analysis in water distribution network design', *Civil Engineering Systems*, 4(4), 175-184.

[24] Goulter, I. C., and A. V. Coals, (1986). 'Quantitative approaches to reliability assessment in pipe networks', *Journal of Transportation Engineering*, 112(3), 287-301.

[25] Goulter, I. C., J. Davidson, and P. Jacobs, (1990). 'Predicting water main breakage rates', ASCE Water Resources Symposium "Infrastructure: Needs Economics and Financing", Fort Worth, Texas

[26] Goulter, I. C. and A. Kazemi, (1988). 'Spatial and temporal groupings of water main pipe breakage in Winnipeg', *Canadian Journal of Civil Engineering*, 15(1), 91-97.

[27] Goulter, I. C. and A. Kazemi, (1989). 'Analysis of water distribution pipe failure in Winnipeg, Canada', *Journal of Transportation Engineering*, 115(2), 95-111.

[28] Goulter, I. C., Bernard M. Lussier, and David R. Morgan, (1986). 'Implications of head loss path choice in the optimization of water distribution networks', *Water Resources Research*, 22(5), 819-822.

[29] Goulter, I. C., David Morgan, (1984). 'Discussion of "Obtaining the layout of water distribution systems" by Rowell and Barnes', *Journal of Hydraulic Engineering*, 110(1), 67-68.

[30] Goulter, I. C. and D. R. Morgan, (1985). 'An integrated approach to the layout and design of water distribution networks', *Civil Engineering Systems*, 2(2), 104-113.

[31] Jacobs, P. and I. Goulter, (1988). 'Evaluation of methods for decomposition of water distribution networks for reliability analysis', *Civil Engineering Systems*, 5(2), 58-64.

[32] Jacobs, P. and I. Goulter, (1989). 'Optimization of redundancy in water distribution networks using graph theoretic approaches', *Engineering Optimization*, vol. 15(1), 71-82.

[33] Jacobs, P. and I. Goulter,'Estimation of maximum cutset size for water network failure', (1990). *Journal or Water Resources Planning and Management*, 117(5), 588-605.

[34] Jeppson, Roland W. and Allen L. Davis, (1975). 'Pressure reducing valves in pipe network analysis', *Journal of the Hydraulics Division, ASCE*, 102(HY7), 987-1001.

[35] Jeppson, Roland W., and Alireza Tavallee, (1975). 'Pumps and reservoirs in networks by linear theory', *Journal of the Hydraulics Division, ASCE*, 101(HY3), 576-580.

[36] Johnson, Norman L. and Samuel Kotz, *Distributions in Statistics: Continuous Multivariate Distributions.* John Wiley and Sons, New York, 1970, pp. 333.

[37] Karmeli, David, Y. Gadish, and S. Meyers, (1968). 'Design of optimal water distribution networks', *Journal of the Pipeline Division, ASCE*, 94(PL1),1-10.

[38] Kettler, A. J. and I. C. Goulter, (1985). 'An analysis of pipe breakage in urban water distribution networks', *Canadian Journal of Civil Engineering*, 12(2), 286-290.

[39] Le, Khiem V. and Victor O. K. Li, (1990). 'Modeling and analysis of systems with multimode components and dependent failures', *IEEE Transactions on Reliability*, 38(1), 68-75.

[40] Mays, L. W., H. G. Wenzel, and J. C. Liebman, (1976). 'Model for layout of sewer systems', *Journal of the Water Resources Planning and Management Division, ASCE*, 102(WR2), 384-405.

[41] Meyer, Paul L., (1970). *Introductory Probability and Statistical Applications.* Addison Wesley, New York, pp. 367.

[42] Morgan, D. R. and I. C. Goulter, (1985). 'Optimal urban water distribution design', *Water Resources Research*, 21(5), 642- 652.

[43] Nel, Louis D. and Charles J. Colbourn, (1990). 'Combining Monte Carlo estimates and bounds for network reliability', *Networks*, 20(3), 277-298.

[44] Neter, John, William Wasserman, and Michael H. Kutner, (1985). *Applied Linear Statistical Methods*, Irwin, Homewood, Illinois, pp. 1127, 2nd edition.

[45] O'Day, K., (1982)."Organizing and analysing leak and break data for making main replacement decisions". *Journal of the American Waterworks Association*, 74(11), 588-594.

[46] Quimpo, Rafael G. and Uzair M. Shamsi, (1988). 'Network reliability for water distribution system management', Fifth IAHR International Symposium on Stochastic Hydraulics, University of Birmingham, Birmingham, U. K.

[47] Quimpo, Rafael G. and Uzair M. Shamsi, (1991). 'Reliability-based distribution system maintenance', *Journal of Water Resources Planning and Management*, 117(3), 321-339.

[48] Quindry, Gerald E., E. Downey Brill, and Jon C. Liebman, (1981). 'Optimization of looped water distribution systems', *Journal of Environmental Engineering Division, ASCE*, 107(EE4), 665-679.

[49] Provan, J. S. and M. O. Ball, (1983). 'The complexity of counting cuts and of computing the probability that a graph is connected', *SIAM Journal of Computations*, 12(4), 777-788.

[50] Rosenthal, Armon, (1981). 'Series-parallel reduction for difficult measures of network reliability', *Networks*, 11(4), 323-334.

[51] Rowell, William F. and J. Wesley Barnes, 'Obtaining the layout of water distribution systems', (1982), *Journal of the Hydraulics Division, ASCE*, 108(HY1), 137-148.

[52] Satyanaraya, A. and Mark K. Chang, (1983). 'Network reliability and the factoring theorem', *Networks*, 13(1), 107-120.

[53] Satyanarayana, A. and Jane N. Hagstrom, (1981), 'Combinatorial properties of directed graphs useful in computing network reliability', *Networks*, 11(4), 357-366.

[54] Shamir, Uri, (1974). 'Optimal design and operation of water distribution systems', *Water Resources Research*, 10(1), 27-36.

[55] Shamsi, Uzair M., (1990). 'Computerized evaluation of water-supply reliability',*IEEE Transactions on Reliability*, 39(1), 35-41.

[56] Su, Yu-Chun, Larry W. Mays, Ning Duan, and Kevin E. Lansey, (1987). 'Reliability-based optimization model for water distribution systems', *Journal of Hydraulic Engineering, ASCE*, 114(12), 1539-1556.

[57] Templeman, A. B. and D. F. Yates, (1984). 'Mathematical similiarities in engineering network analysis', *Civil Engineering Systems*, 1(1), 114-122.

[58] Wagner, J. M., U. Shamir, and D. H. Marks, (1988a). 'Water distribution reliability: analytical methods', *Journal of Water Resources Planning and Management, ASCE*, 114(3), 253-275.

[59] Wagner, J. M., U. Shamir, and D. H. Marks, (1988b). 'Water distribution reliability: simulation methods', *Journal of Water Resources Planning and Management, ASCE*, 114(3), 276-294.

[60] Walski, Thomas M., E. Downey Brill, Jr., Johannes Gessler, Ian C. Goulter, Roland M. Jeppson, Kevin Lansey, Han-Lin Lee, Jon C. Liebman, Larry Mays, David R. Morgan, and Lindell Ormsbee, (1987). 'Battle of the network models:

epilogue', *Journal of Water Resources Planning and Management*, 113(2), 191-203.

[61] Watanatada, T., (1973). 'Least cost design of water distribution systems', *Journal of the Hydraulic Division, ASCE*, 99(HY9), 1497-1513.

[62] Wood, Don J., (1980). *KYPIPES*, Engineering Continuing Education, University of Kentucky.

[63] Wood, Don J. and Carl O. A. Charles, (1972). 'Hydraulic network analysis using linear theory', *Journal of the Hydraulics Division, ASCE*, 98(HY7), 1157-1170.

[64] Wu, Sue-Jen and R. G. Quimpo, (1992). 'Predictive model and reliability analysis for water distribution systems', *Proceedings of the Sixth IAHR International Symposium on Stochastic Hydraulics*, Taipei, 163-170.

[65] Zemel, E., (1982). 'Polynomial algorithms for estimating network reliability', *Networks*, 12(4), 439-452.

Appendix A

Data Transformation and Verification

A.1 Description of the Pipe Break and Pipe Inventory Data

Two types of data are used in the model presented in Chapter 5. These data types are pipe failure history and pipe inventory data. As noted previously the pipe failure history data are contained in a data base gathered by the Waterworks, Waste and Disposal Department of the City of Winnipeg. This is a computerized data base that contains one failure record for each pipe break that occurred in the years 1975 through 1989. In addition, the data base contains location records for a large number of physical features found in the city. Over 30,000 breaks are recorded in this data base.

The location records in this data base contain the x-y coordinates of every significant feature of the road system within the City of Winnipeg when the road system was digitized in 1975. Significant features of the road system include intersections with other roads, intersections with physical features such as rivers and railroads, and places where roads change direction. These location records are stored alphabetically by street name. It should be noted that the location records do not contain any pipe inventory data describing the pipe characteristics at the location. Of particular consequence for this study is the fact that location records do not contain pipe diameter, material, or installation date.

The failure records are interspersed among the location records and contain x-y coordinates of the break location which are accurate to the nearest meter. These break records contain failure date, pipe installation date, pipe diameter, pipe material, and other parameters which are used by this study.

This data base proved to be extremely valuable for the type of analysis described in Chapter 5, but has a number of weaknesses which should be recognized. One problem is that the data base does not contain start and end times for failure events. This shortcoming makes it impossible to be certain whether or not failure events are simultaneous. It is possible for two breaks to occur on the same day, but for the first break to be repaired before the second break occurs. On the other hand, it is possible for a break to start on one day and not be repaired before the next day. This situation is almost certain for breaks which start late in the evening, but can also occur if repair crews are already busy when a break occurs. If the work crews are busy it is possible that breaks with less serious consequences would be left while more serious breaks are repaired. Due to this limitation on the accuracy of the data, breaks which occur on the same day are assumed to be simultaneous and all breaks are assumed to be fixed on the day of occurrence.

A second weakness of the data base is that it contains the location of street centre lines rather than the location of the water mains themselves. As water mains are typically laid beside a roadway rather than under the centre line, this weakness introduces a small locational error. For the purposes of this study, this error was ignored.

A third weakness of this data base is that pipe inventory data are stored in the break records. As a result, if there is no break record for a street, there is no pipe inventory data for that street. From the point of view of this study the effect of this weakness in the data base organization results in there being no location records for streets which are not serviced by water mains. Furthermore, streets not serviced by water mains are stored in exactly the same way as serviced streets on which no breaks have occurred in the period of record. As a result it is not possible to distinguish between streets with no pipes and streets with no breaks. This is not an insignificant problem as approximately 1200, or slightly more than one third, of the approximately 3500 streets in the City of Winnipeg have never had a break and possibly have no watermains.

A fourth weakness of the data base is that the information in it is gathered, but not directly used, by operational personnel. Under these conditions, verification of the data, contained in the data base, is extremely difficult. As a result even when failure records have been added for a street there may be errors in pipe characteristics in the data base.

Assuming that the break data base without support from other data bases, is used the effects, of the third and fourth of these weaknesses are that no pipe inventory data exists for approximately one third of all streets in the City of Winnipeg. It was therefore necessary to transform, augment, and verify the break data base before it could be used in the study.

The first step in transforming the data into a more useful form is conceptual. In the City of Winnipeg break data base information is stored as a series of nodes. Two types of nodes are used: locational nodes and failure nodes. The locational nodes contain some information about the type of physical feature that they represent. For example, nodes that represent a rail road are coded, 'R*', indicating their function, and nodes representing streets can be similarly distinguished from other types of nodes. The locational nodes do not, however, contain any information concerning characteristics of pipe with which they are related. The addition of pipe characteristics to locational nodes leads to a data organization that is awkward and difficult to use because each node is the boundary between pipe segments. If pipe characteristics differ among the pipe segments attached to a node, it is necessary to include both sets of pipe characteristics in the node record and to indicate which characteristics apply to each pipe segment. In addition, if pipe characteristics are applied to node records, then it is also necessary to assume that pipe characteristics remain constant between adjoining nodes.

A more simple method of data organization is to create pipe segment records. In this situation pipe characteristics can be attached to pipe segment records. In addition, by using variable length pipe segments, the pipe segments can be set up so that only one set of pipe characteristics is necessary for each pipe segment. The choice of an appropriate unit for pipe segments is facilitated by the usual practice, for urban water distribution systems, of laying water mains for a minimum length of a full city block at a time. As a result of this practice, pipe characteristics can be reasonably expected to remain constant over the length of a city block.

Storage of pipe characteristics using pipe segment records also facilitates the s-torage of failure records. By definition, pipe failures occur at a particular location in a pipe segment. As a result, failure can always be assigned to a particular pipe segment by its physical location. This relationship makes it possible to store pipe failure records for a particular pipe segment record as a group and to access these failure records using a pointer from the pipe segment record. This storage method has the advantage that the pipe characteristics can be stored on the pipe segment records. Once such characteristics have been added to the pipe segment records, it is then unnecessary to store them on the pipe failure records. A further practical benefit of this storage method is that operational personnel do not have to capture pipe characteristics for every break that occurs as this information can be retrieved from the pipe inventory records. Based on these considerations, the City of Winnipeg break data base was transformed to the storage format described above using a PL1 program.

The second step in transforming the data base is to add missing data and to verify the existing data. Since no computerized pipe characteristics data existed for a large proportion of the City of Winnipeg network it was necessary to add the pipe characteristics to the data base from paper records. Adding these data for the whole City of Winnipeg, would require a large amount of repetitive effort. To limit the amount of repetitive effort, the scope of this study was limited to one (of six) operating districts in the City of Winnipeg, District 4. The amount of data included in the study area is still relatively large. The study includes 394 km of water mains, 3853 breaks, and 3126 pipe segments, over 15 years of record.

The obvious difficulty of this task is further compounded by the characteristics of the paper records used to store water main data and the need to capture the geographical position of data components. The complete pipe inventory is stored on a variety of media including: large scale drawings, kardex, microfiche, and word processor files. Although the data in the pipe inventory are organized geographically, for the most part explicit coordinates are not identified with objects contained in the drawings. Where the drawings are to scale, extracting coordinates is a cumbersome process at best. In many cases, however, the data are stored as 'not to scale' sketches. For this 'not to scale' data, it is not possible to measure coordinates directly.

To aid in the data capture task computer graphics techniques were used. These techniques were implemented by displaying a map of the street grid in a part of the area to be processed using the AUTOCAD program. Several small programs or commands, written in LISP, allowed the manipulation of the elements in the street network.

These computer graphics techniques have a number of advantages. When comparing a street layout for a small area to a map of the same area, it is easy to determine where a particular block is located in the street layout, even if street names are not displayed. The shape, orientation, size, and relative position with respect to any landmarks displayed, all provide easy-to-use visual clues. As a backup to these visual clues, a command in LISP was written to display the street name for a selected city block. To identify a city block it is necessay only to specify the street names at each end. These street names at each end of the block must be identified precisely in the data base. The location of the city block is then defined with certainty and, because AUTOCAD stores the coordinates of the block, the coordinates are also identified with certainty.

The above method makes it possible to augment and verify the data contained in the break data base with information from both 'to-scale' maps and other 'not-to-scale' information sources such as kardex and microfiche. Implementing this method required the development of computer graphics methodology. The computer graphics techniques required to implement this method is described in the next section.

A.2 Techniques to Create the Pipe Inventory

The computer system developed to verify the pipe inventory used both a large mainframe and personal computers. The methodology included computer programs written in LISP, C, and PL1. The steps followed in this procedure are shown schematically in Fig. A.1 and are as follows.

1. capture the physical position of the pipe in the water distribution network
2. add street names to each link in the network
3. load the pipe characteristics from the break records
4. use graphical techniques to verify pipe characteristics loaded from the break data and to add data where they are missing. The results of this step are stored in intermediary records.
5. use the intermediate transactions generated in step 4 to modify and/or verify the data for the network generated in steps 1, 2, and 3.

The location of the centre lines of almost all streets in existence in the City of Winnipeg is contained in the location records in the City of Winnipeg break data base. (Not all streets are included in the break data base, however, because the data base was created in 1975. Streets have been added to the network since 1975. If the water mains on the newly added streets have not experienced a break, the streets have probably not been added to the break data base.)

A copy of the City of Winnipeg break data base is stored on the mainframe computer at the University of Manitoba. A PL1 program was developed to extract information from this data base. This program is identified as 'Extract from Mainframe' in Fig. A.1 and produces the Link File L4F007. In this file one record is created for each city block or pipe segment with each record of file L4F007 containing street names and, where available, pipe characteristics. To improve the response time provided by AUTOCAD, and to facilitate transfer of the data from the main frame computer, the study area was broken into three rectangular areas with one file being extracted for each area.

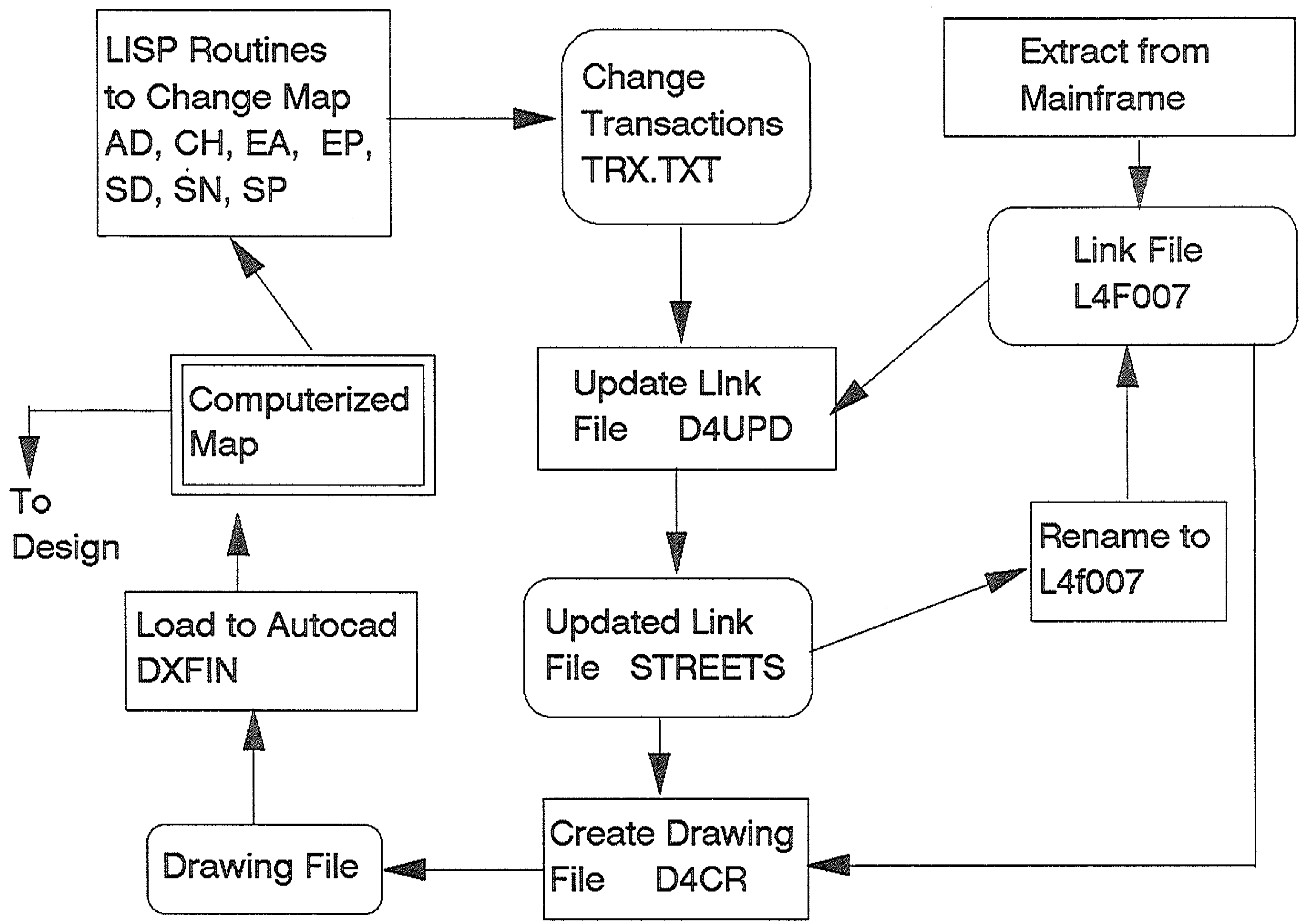


Figure A.1 Schematic of Computerized Map Correction System

The program D4CR (written in C) shown in Fig. A.1 was used to change the format of each of these files to the standard AUTOCAD DXF format. This program also added codes to the file to vary the color of pipe segments to differientiate between the different types of information available for the pipe in the segements when they were displayed. Pipe segments with missing, or incomplete, pipe characteristics data were colored red. Pipe segments containing complete pipe characteristics were colored green. The use of different colors makes it easy to determine which pipe segments are missing data.

Once the files were reformatted into DXF format, the built-in AUTOCAD command DXFIN was used to load them to the AUTOCAD workspace. The files could then be displayed as a computerized map. The data contained in this map are still in a relatively raw form because they have not been augmented or verified. Fig. A.2 shows a section of a computerized map at this stage of processing. For ease of printing, different line formats, rather than colors, are used to distinguish pipe segments that have pipe characteristic data from those that do not.

Before these raw data can be used as a pipe inventory, a number of processing tasks must be performed. These tasks include

1. The removal of pipe segments not in District 4. District 4 does not have rectangular boundaries. The use of rectangular boundaries, in the extraction process, resulted in the initial inclusion of pipe segments from other operating districts

2. The removal of street segments which do not contain mains. Since the City of Winnipeg break data base does not distinguish between pipe segments which have never experienced a break and streets with no water mains, it is necessary to remove the latter manually

3. The addition of pipe segments that were not included during the creation of the City of Winnipeg break data base

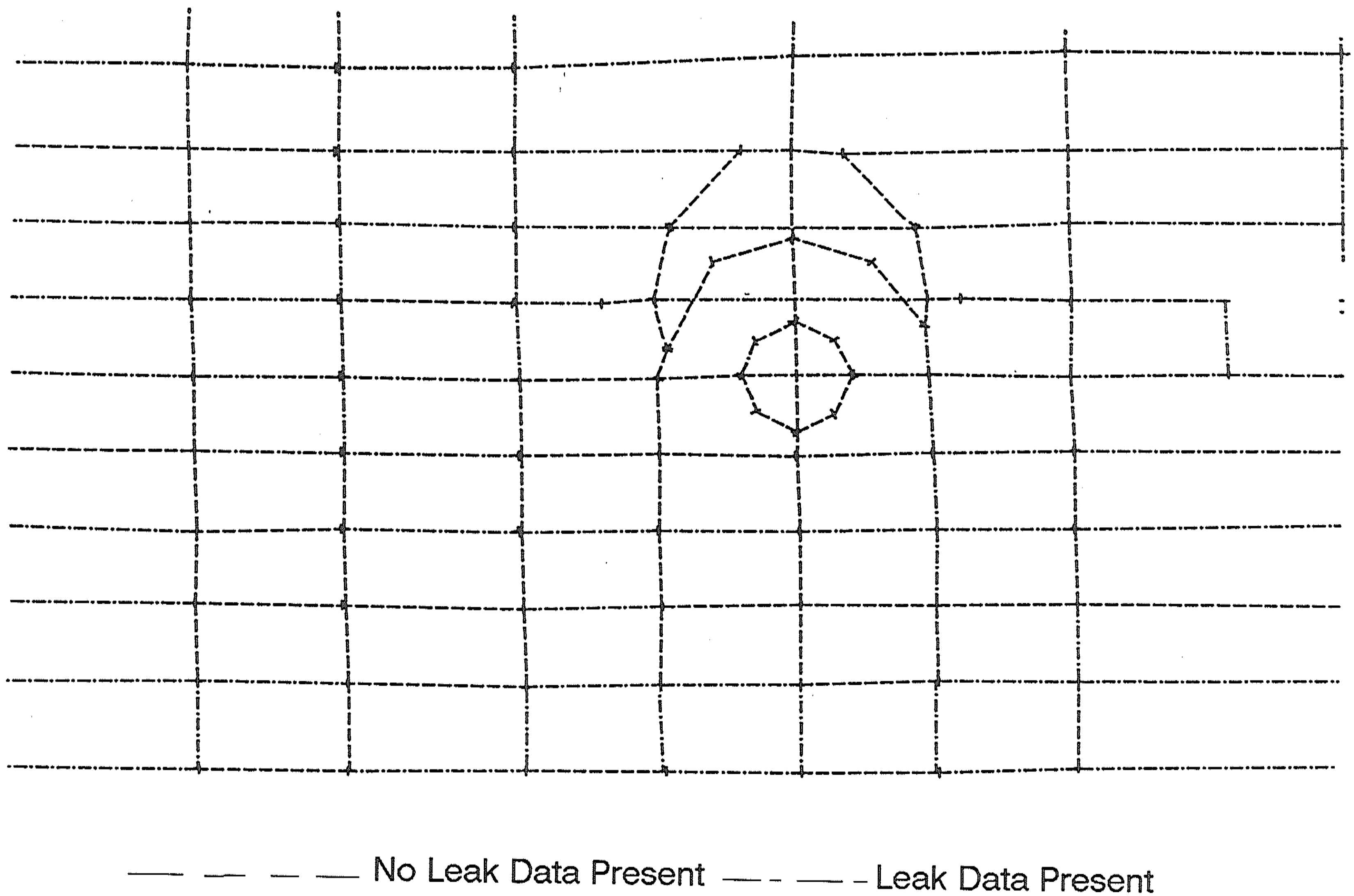


Figure A.2 Segment of Computerized Map Before Validation

4. The modification of pipe segments where pipe renewals have taken place

5. The addition of pipe characteristics where they are missing

6. The verification of pipe characteristics where they exist. For those pipe segments in which the pipe characteristics in the break data base differ from those in the operational data, it was assumed that the operational data were correct. The reason for choosing operational rather than break data when there is a conflict relates to the greater likelihood of the operational data being correct because it is used and reviewed more often (Goulter and Jacobs, 1992).

To facilitate these tasks, a number of LISP programs were written. The LISP language runs as a part of the AUTOCAD computer package and is integrated with AUTOCAD to make use of the computer graphics capabilities of AUTOCAD, in particular, those facilities which allow a user to graphically point to and select an object that is displayed.

LISP programs that make use of this point and select facility were written. The LISP programs make it possible to graphically point to and select a pipe segment. Once the pipe segment is selected the data associated with the pipe segment can be manipulated. These data include street name, pipe characteristics, and x-y coordinates. Finally, 'intermediate' transactions can be written using the LISP programs to record changes to the computerized map. The function of each of these programs and their relationship to each other is listed in Table A.1 and discussed in detail below. For the sake of consistency, when a program generates an intermediate transaction, the program name and the transaction code are are designated by the same terms. For example, the add program is named AD and generates an intermediate transaction with transaction code AD.

Add Program (AD):

Program Name	Program Function
AD	Add a pipe segment to the network
CH	Change pipe characteristic for a pipe segment
EA	Delete a pipe segment from the network
EP	Delete part of a pipe segment from the network
SD	Display street name for a pipe segment
SN	Display street name and pipe characteristics
SP	Split a pipe segment into several smaller segments

Table A.1: LISP Macros and Their Functions

The add program is used to add new pipe segments and creates intermediary add transactions (Transaction type AD in Fig. A.1). Add transactions are necessary if the map does not contain a street and/or where renewal has been performed. The computerized map for the City of Winnipeg is generally complete as originally generated. However, as noted previously it is necessary to add some streets in subdivisions built after 1975, as those streets did not exist when the break system was set up. Since the pipe in such subdivisions is, in general, relatively new, no breaks have occurred on a large proportion of these pipe segments and there has been no reason to add them to the City of Winnipeg break data base.

In cases where renewal work was performed in the period 1975-1989, it was necessary to add links to specify the characteristics of the new pipe segments added. It should be noted that addition of these links has the implication that more than one pipe segment can exist for a given geographical location. It is, however, possible to distinguish in which of these multiple pipe segments a break occurred because the date of installation of the pipe is included in the pipe characteristics. By checking the pipe installation date against the date when a break occurred, it is possible to specify, uniquely, which link record should be used.

Change Program (CH)

The change program is used to correct the pipe characteristics for a pipe segment and generates the CH transaction. This program is used when no pipe characteristics are captured from the City of Winnipeg break database or when the captured characteristics are in error. For the City of Winnipeg the most commonly used sources of data are kardex and scale drawings of the pipe layout. These data sources are organized in a fashion that makes it easy to verify contiguous links for a street. In addition, it is common for pipe characteristics to remain constant over a number of contiguous pipe segments. To take advantage of this pattern an option was included in the Change program to copy pipe characteristics to the current pipe segment from an adjacent pipe segment processed previously. Due to the large number of pipe segments in which a break has not occurred in the period of record this program is by far the most commonly used program and the Change transaction is the most common intermediate transaction.

Erase Whole Link Program(EA)

The Erase Whole Link Program generates transaction type EA which is used to remove a pipe segment from the pipe network. It is necessary to remove links from the network because not all city blocks are served by water mains. In addition, as described previously, when the raw data were extracted from the City of Winnipeg break data base, not all of the pipe segments extracted were from the study area, District 4. This program is used to remove these superfluous links.

Erase Part Link Program(EP)

The Erase Part Link program is used to generate the EP transaction. The EP transaction removes part of a pipe segment from the network. This program is used when a water main is terminated in the middle of a pipe segment. This is a rarely used program, reflecting the relative infrequency of this situation.

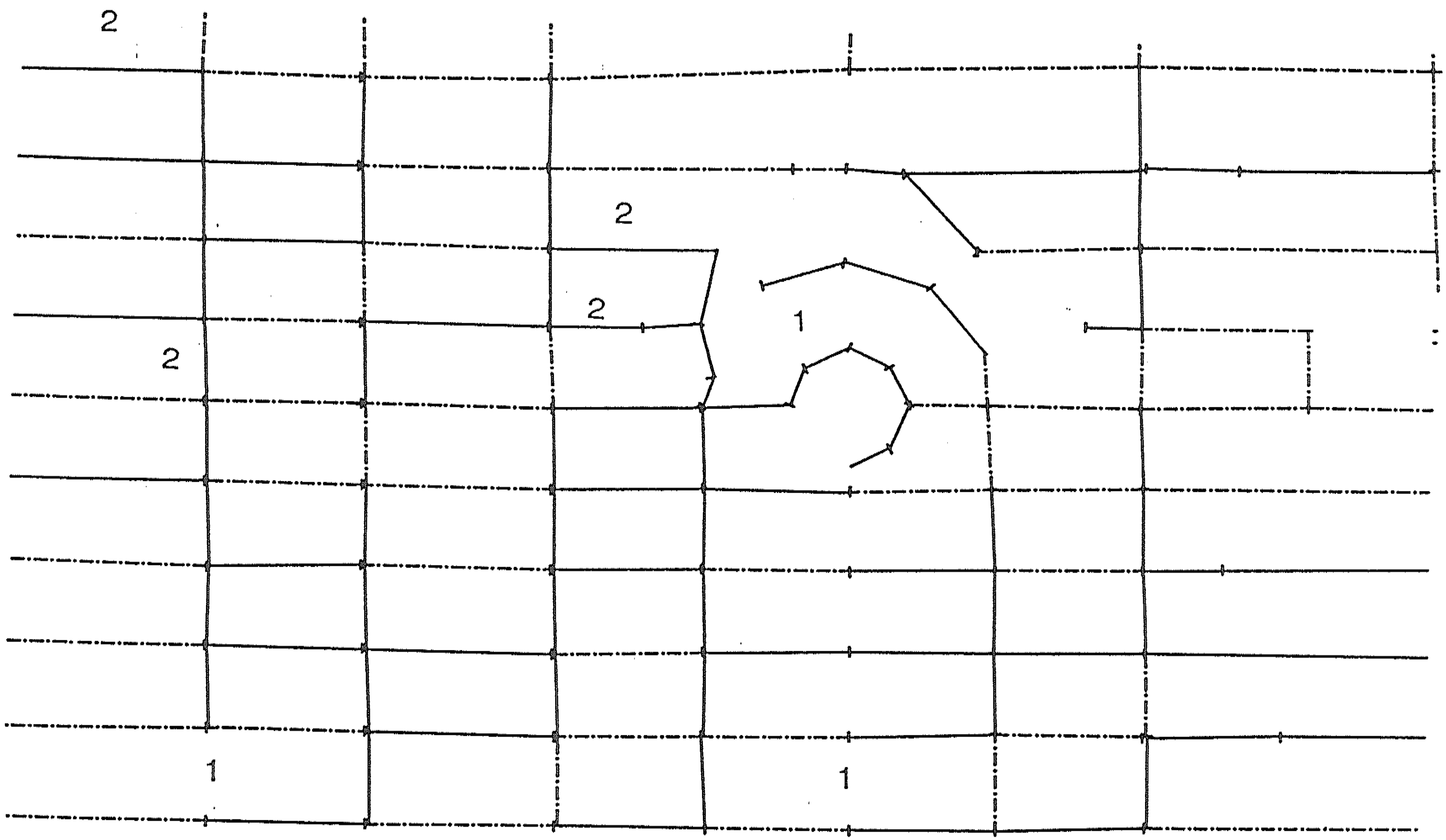
Show Street Name Program (SD)

The Show Street Name program is used select a pipe segment and to display the associated street name. To save re-display and regenerate time, and to make the map less cluttered, the display of street names and pipe characteristics is generally suppressed. This program is used to aid the operator in locating pipe segments on the computer map. This program can be used to positively identify a pipe segment by using it to display the names of streets at the pipe segment ends.

Show Street Characteristics Program (SN)

The Show Street Characteristics program is used to select a pipe segment and to display the street name and pipe characteristics. This program is used to verify pipe characteristics of pipe segments for which the pipe characteristics were captured from the City of Winnipeg break data base. In addition, the street name and pipe characteristics are also stored for use, if the next program run is a Change program. This feature is particularly useful if a street has one pipe segment with correct pipe characteristics in the computerized map and several other pipe segments which should have the pipe characteristics but for which no pipe characteristics were captured. In this case, it is possible to transfer the correct pipe characteristics from pipe segment to pipe segment by running the SD, program followed by multiple runs of the Change program.

The data correction system as outlined in this chapter was run through several iterations until all street segments were labelled as verified. The segment of the computerized map shown in Fig A.2 (uncorrected) is shown after correction in Fig. A.3.



1 -- Leak data but no pipe present 2 -- Leak data present but incorrect

— - — - Leak Data Present ———— Verified

Figure A.3 - The Same Segment of the Computerized Map after Correction

Appendix B

Additional Figures to Estimate Component Failure Probabilities

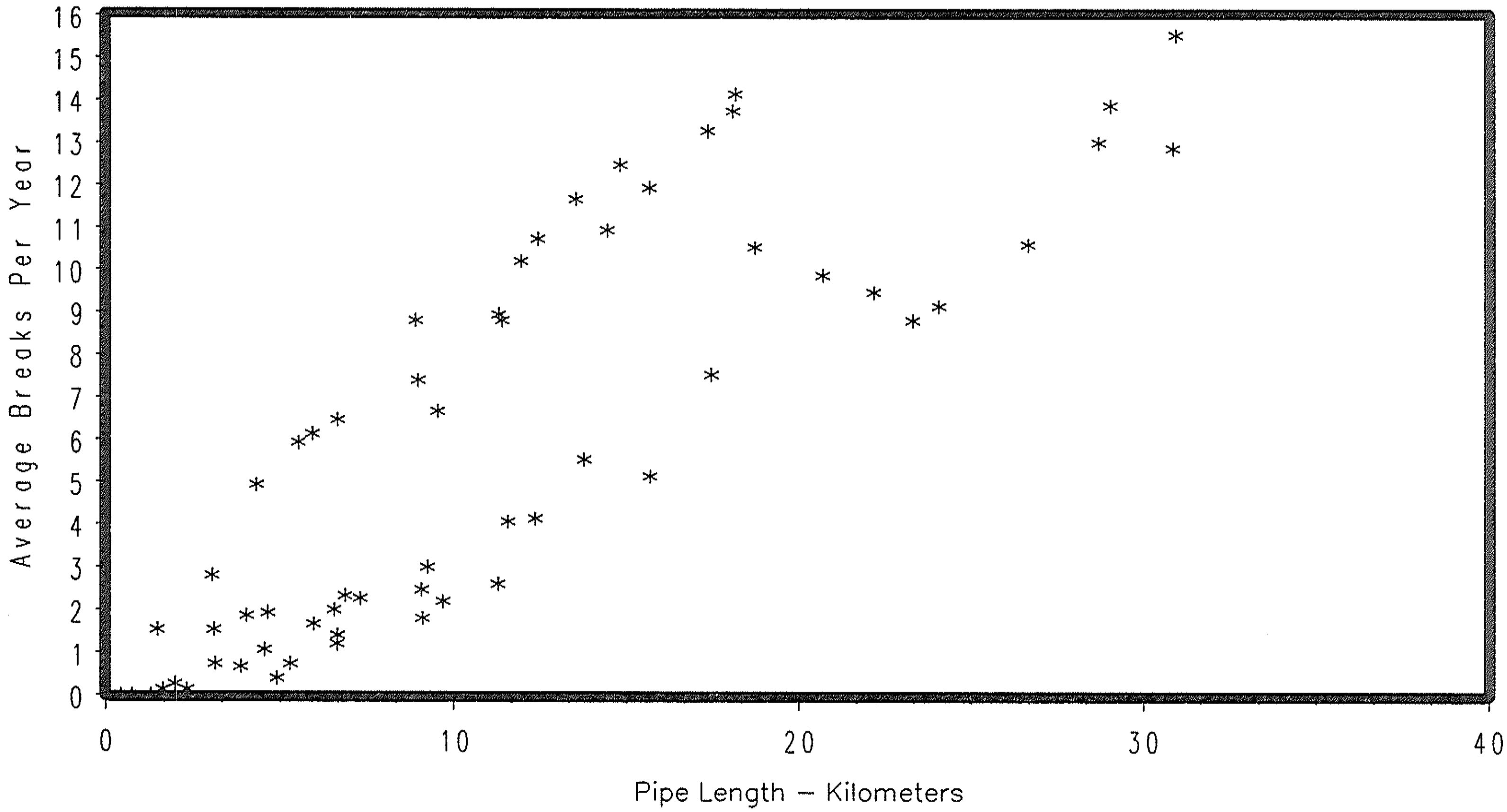


FIGURE B.1 Average number of breaks per km: cast iron pipe, 150mm 0–19 years

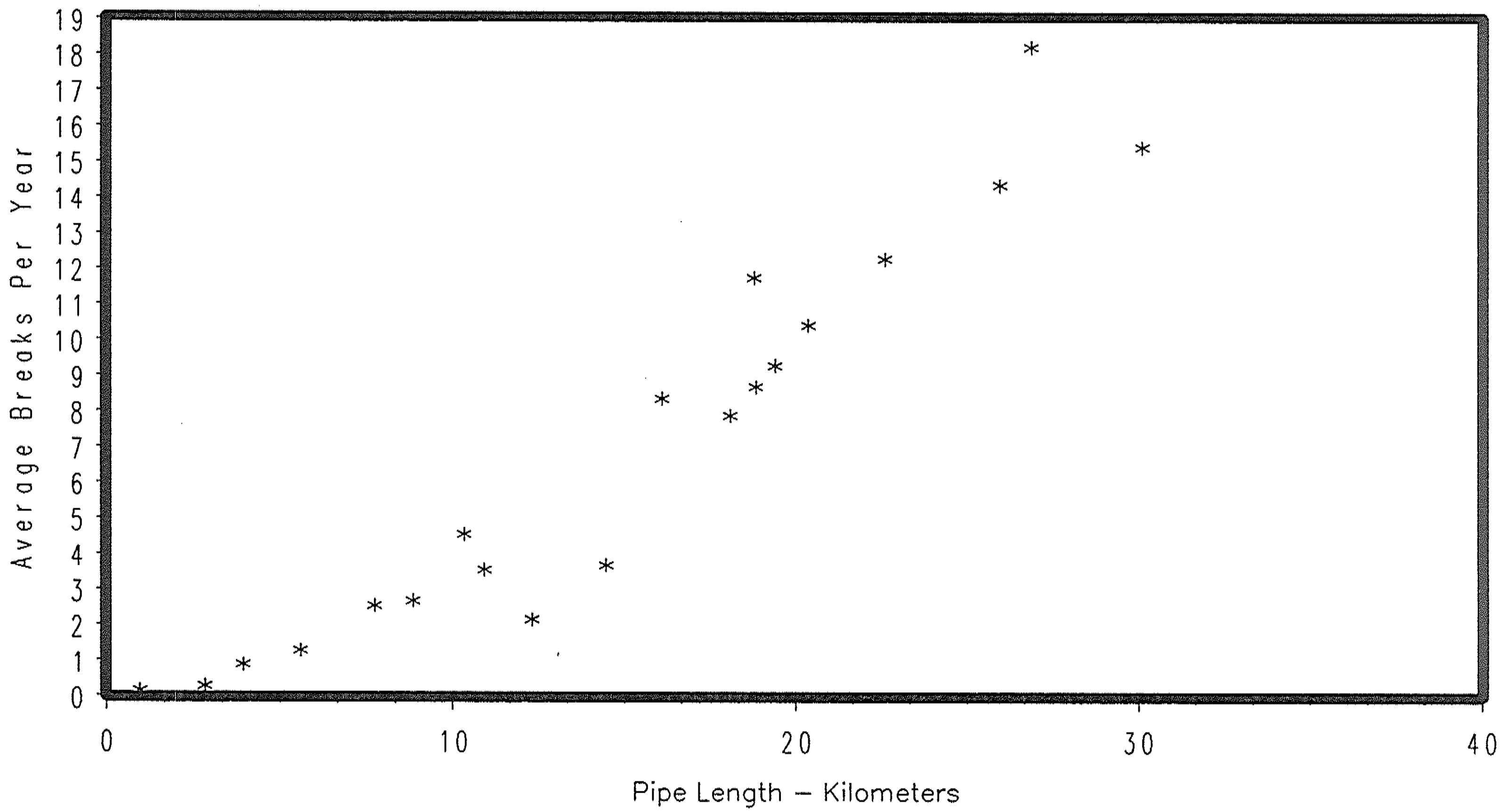


FIGURE B.2 Average number of breaks per km: cast iron pipe, 150mm 20–30 years

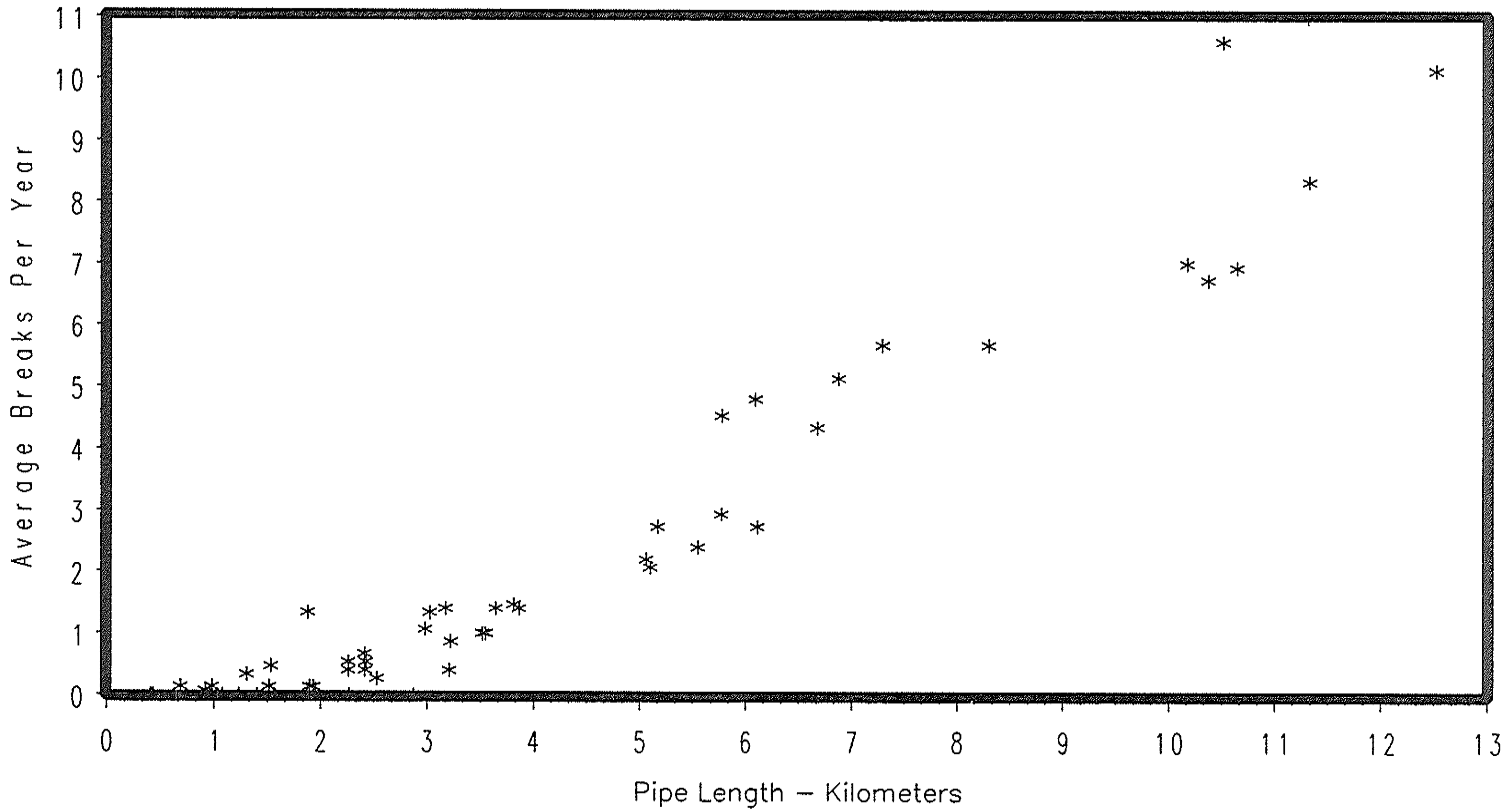


FIGURE B.3 Average number of breaks per km: cast iron pipe, 150mm 31+ years

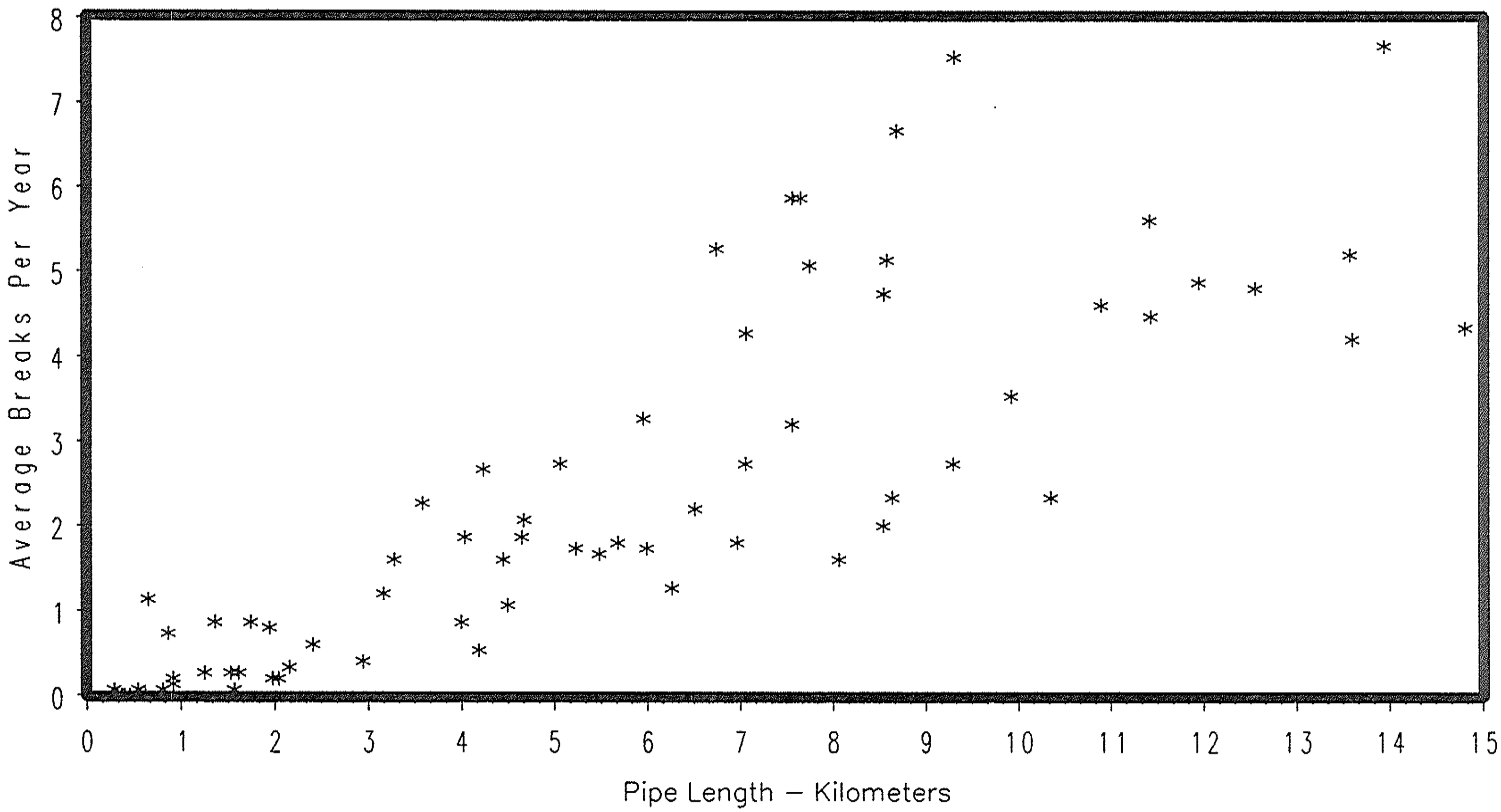


FIGURE B.4 Average number of breaks per km: cast iron pipe, 200mm 0–30 years

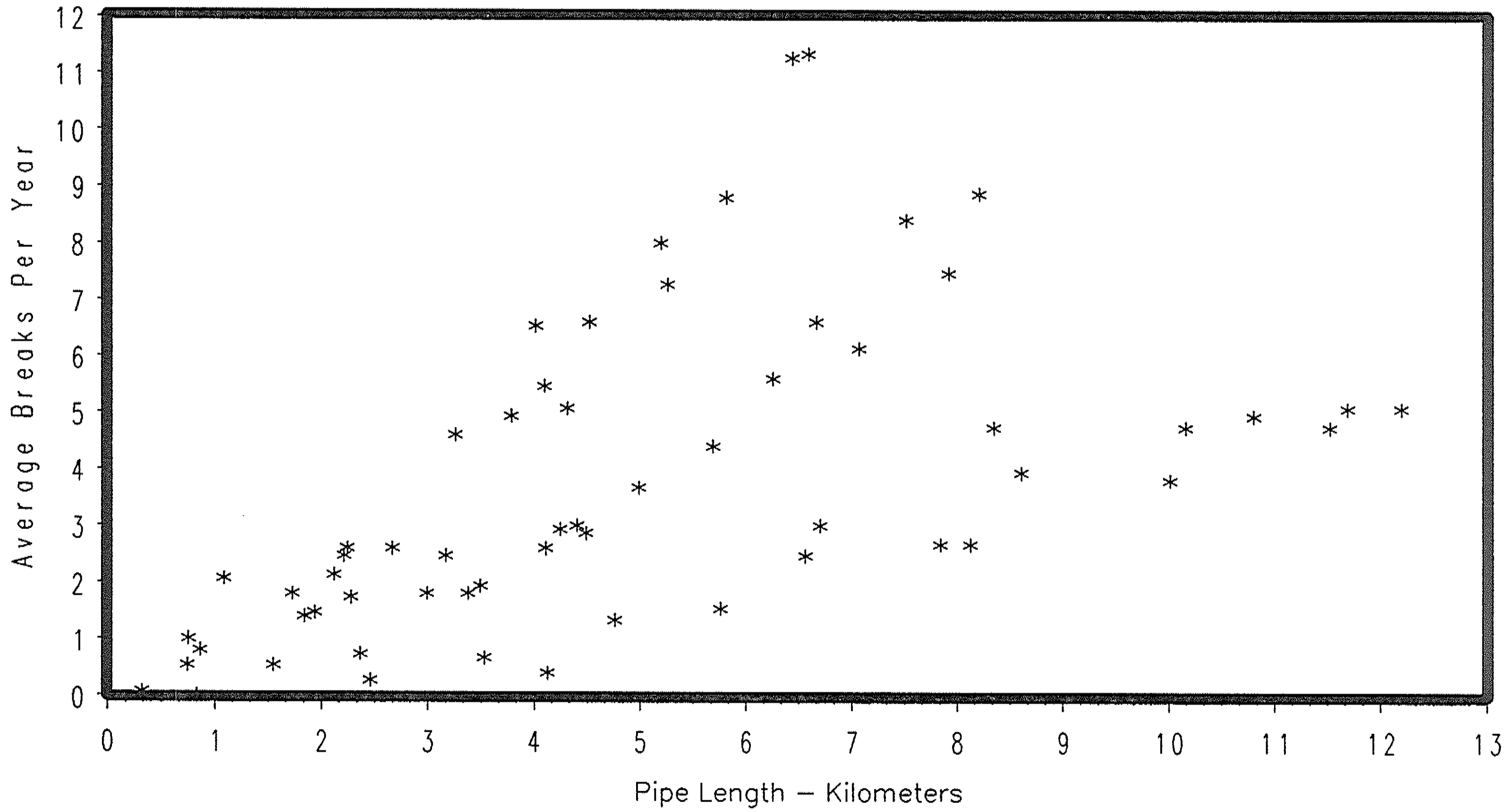


FIGURE B.5 Average number of breaks per km: cast iron pipe, 200mm 31+ years

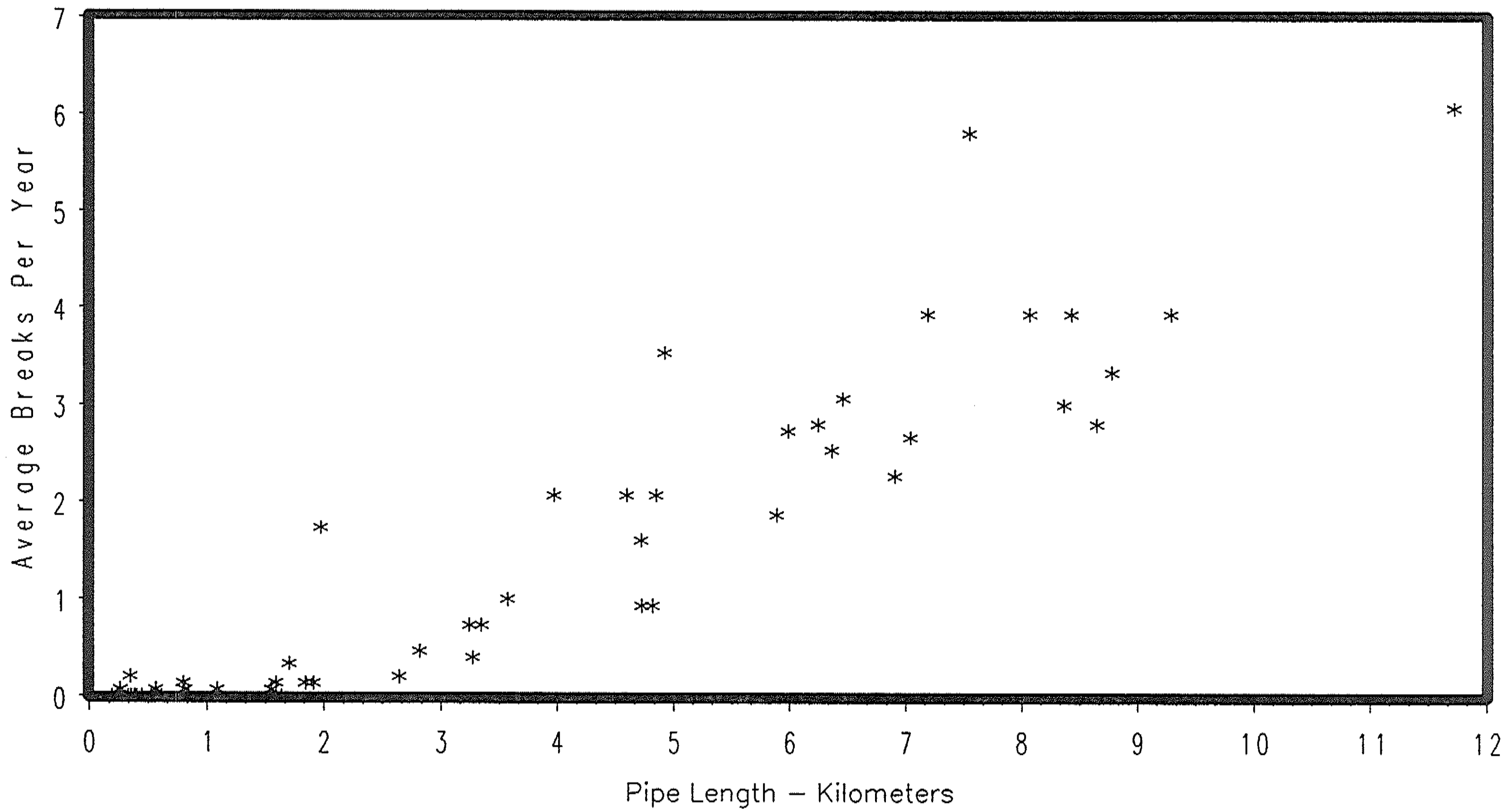


FIGURE B.6 Average number of breaks per km: cast iron pipe, 250mm 0–30 years

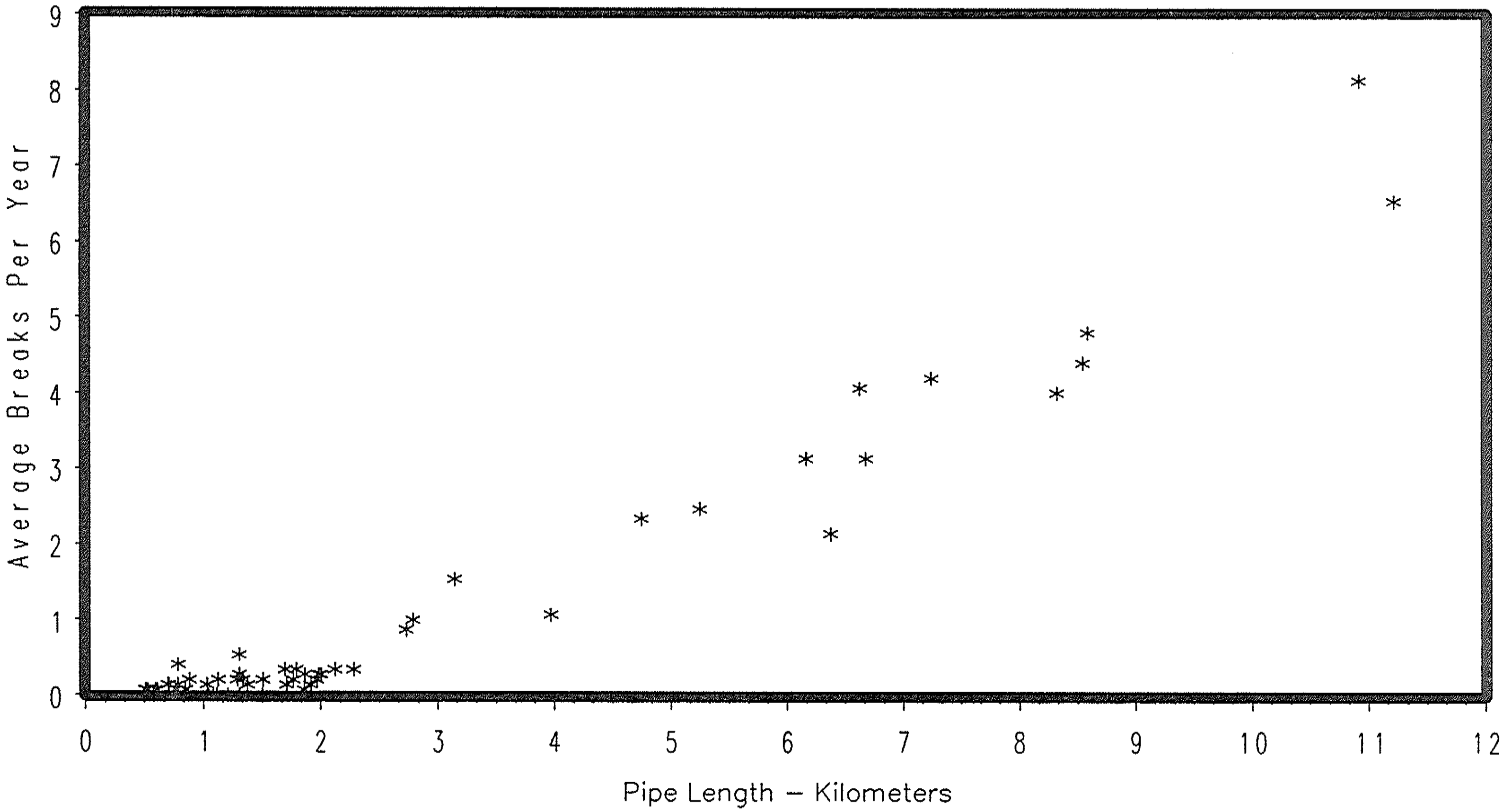


FIGURE B.7 Average number of breaks per km: cast iron pipe, 250mm 31+ years

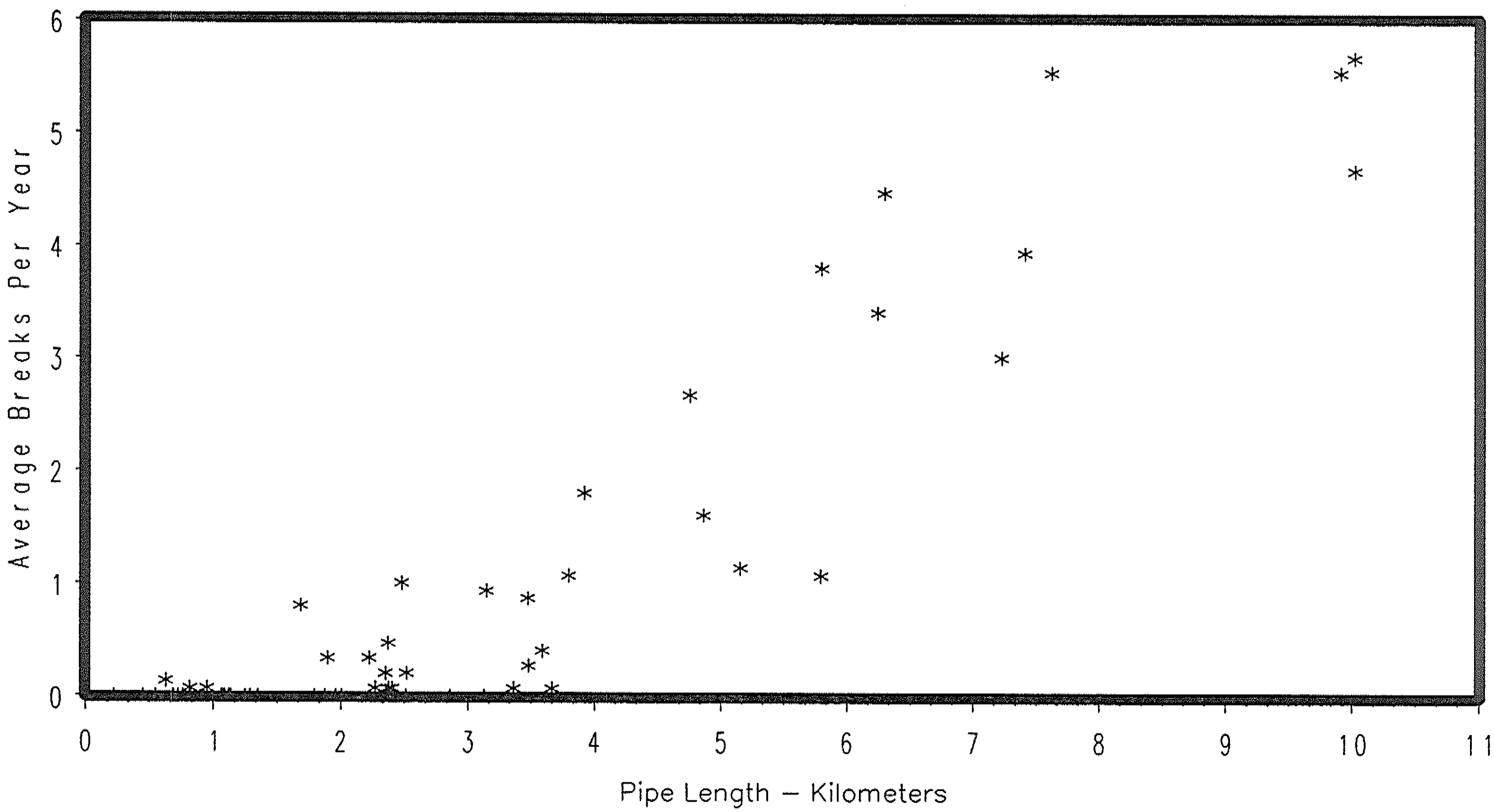


FIGURE B.8 Average number of breaks per km: cast iron pipe, 300mm 0–30 years

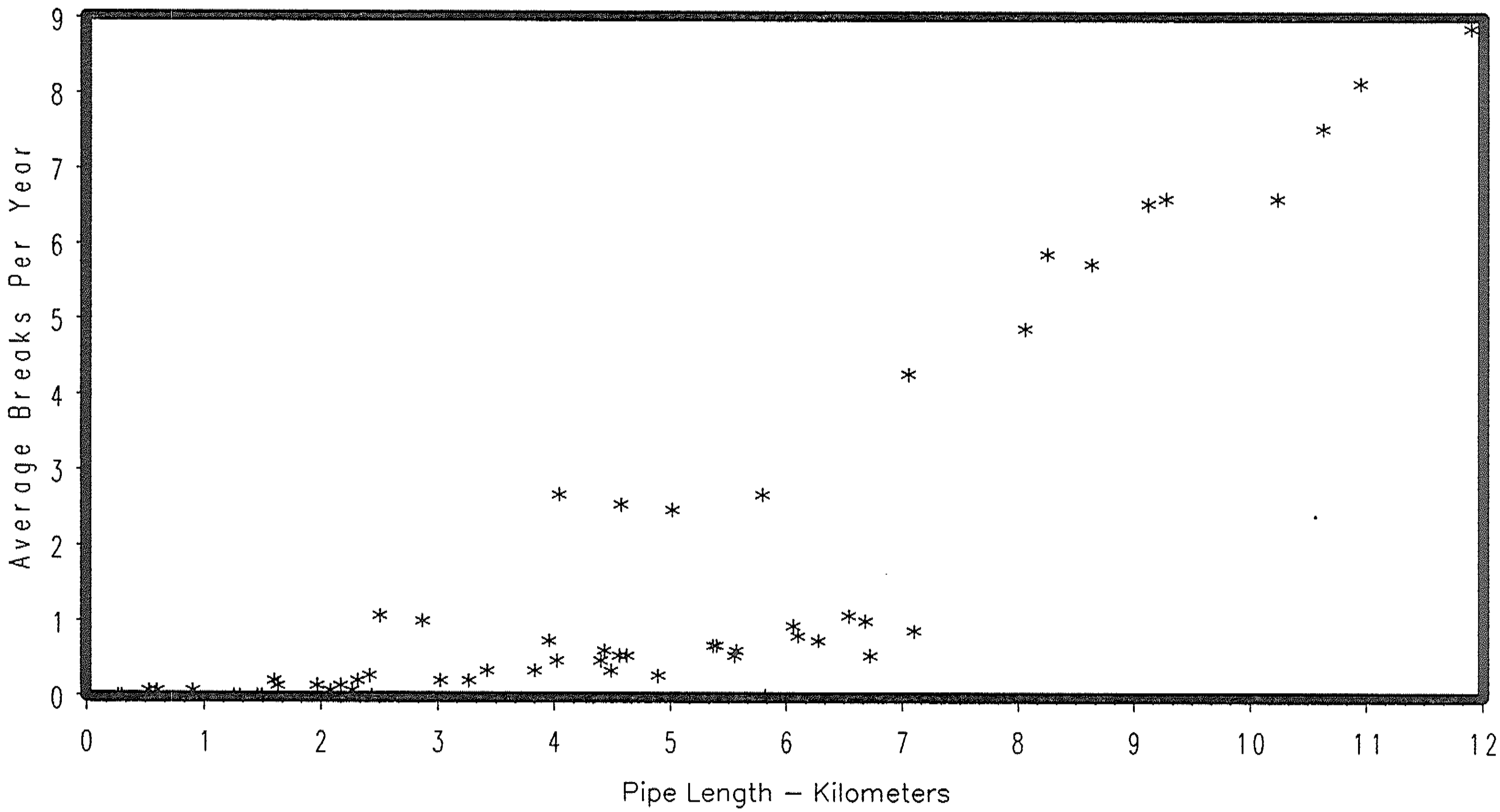


FIGURE B.9 Average number of breaks per km: cast iron pipe, 300mm 31+ years

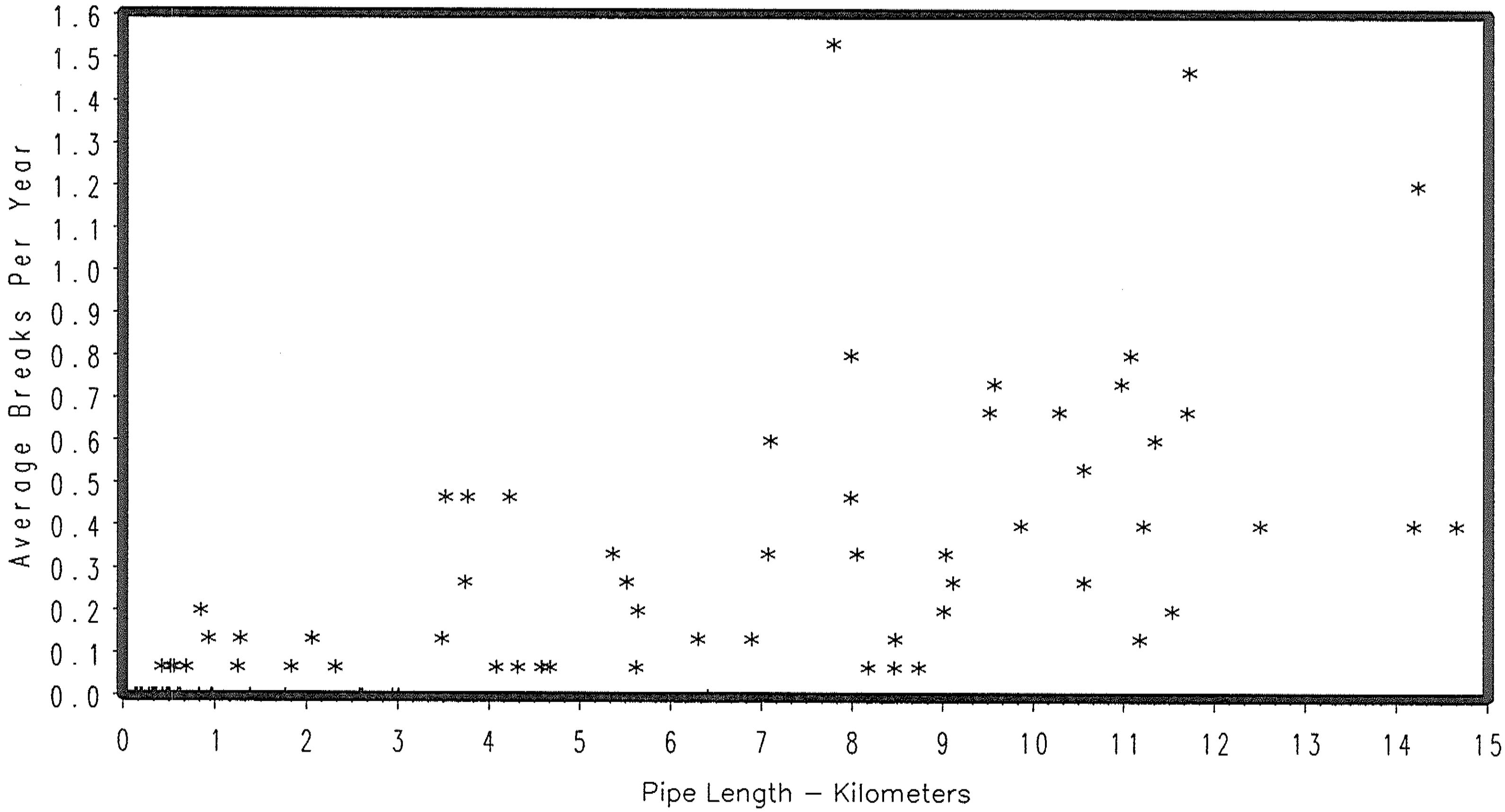


FIGURE B.10 Average number of breaks per km: asbestos cement pipe,150mm 0–30 years

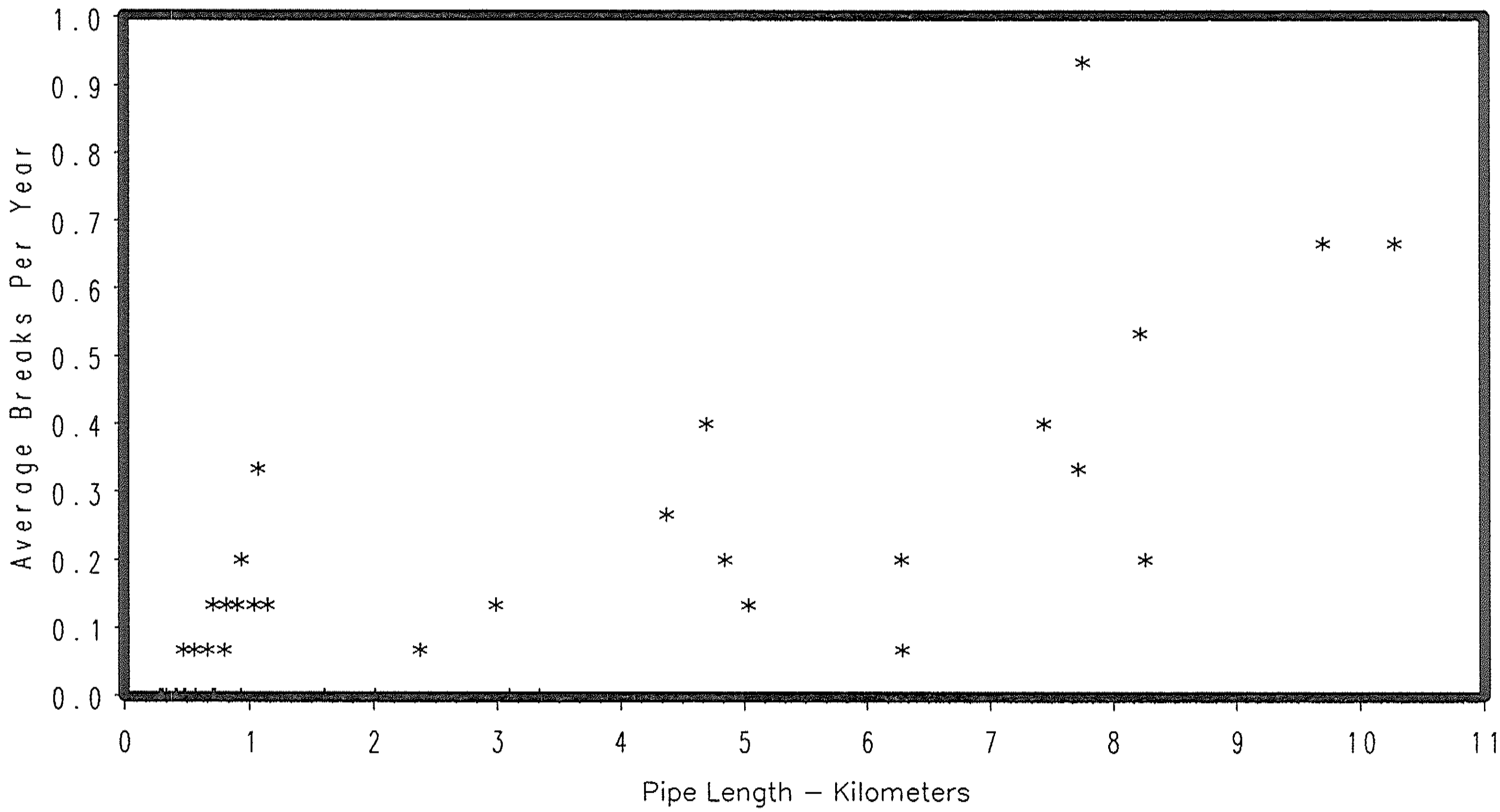


FIGURE B.11 Average number of breaks per km: asbestos cement, 150mm 31+ years

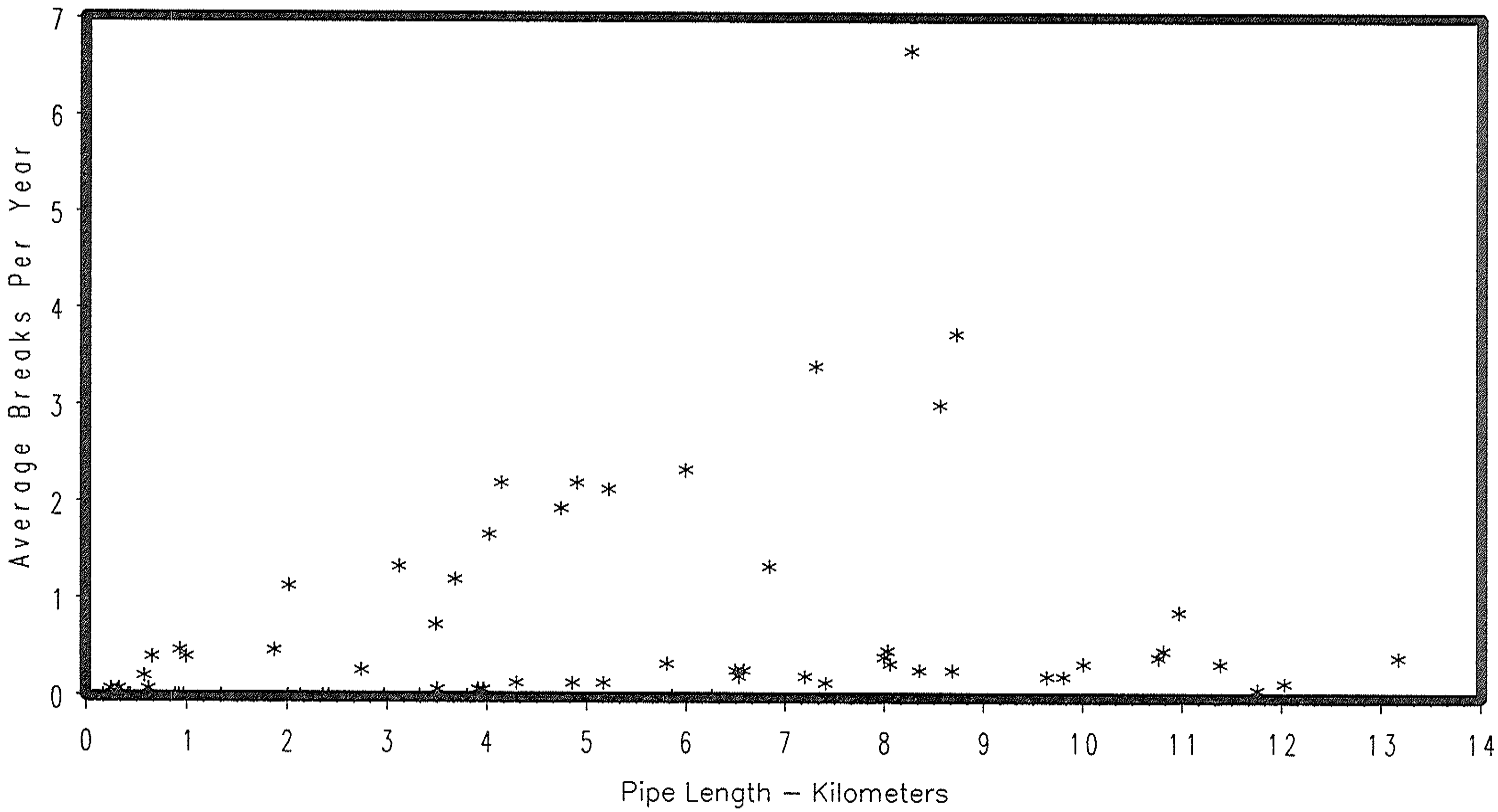


FIGURE B.12 Average number of breaks per km: asbestos cement pipe, 200mm 0–30 years

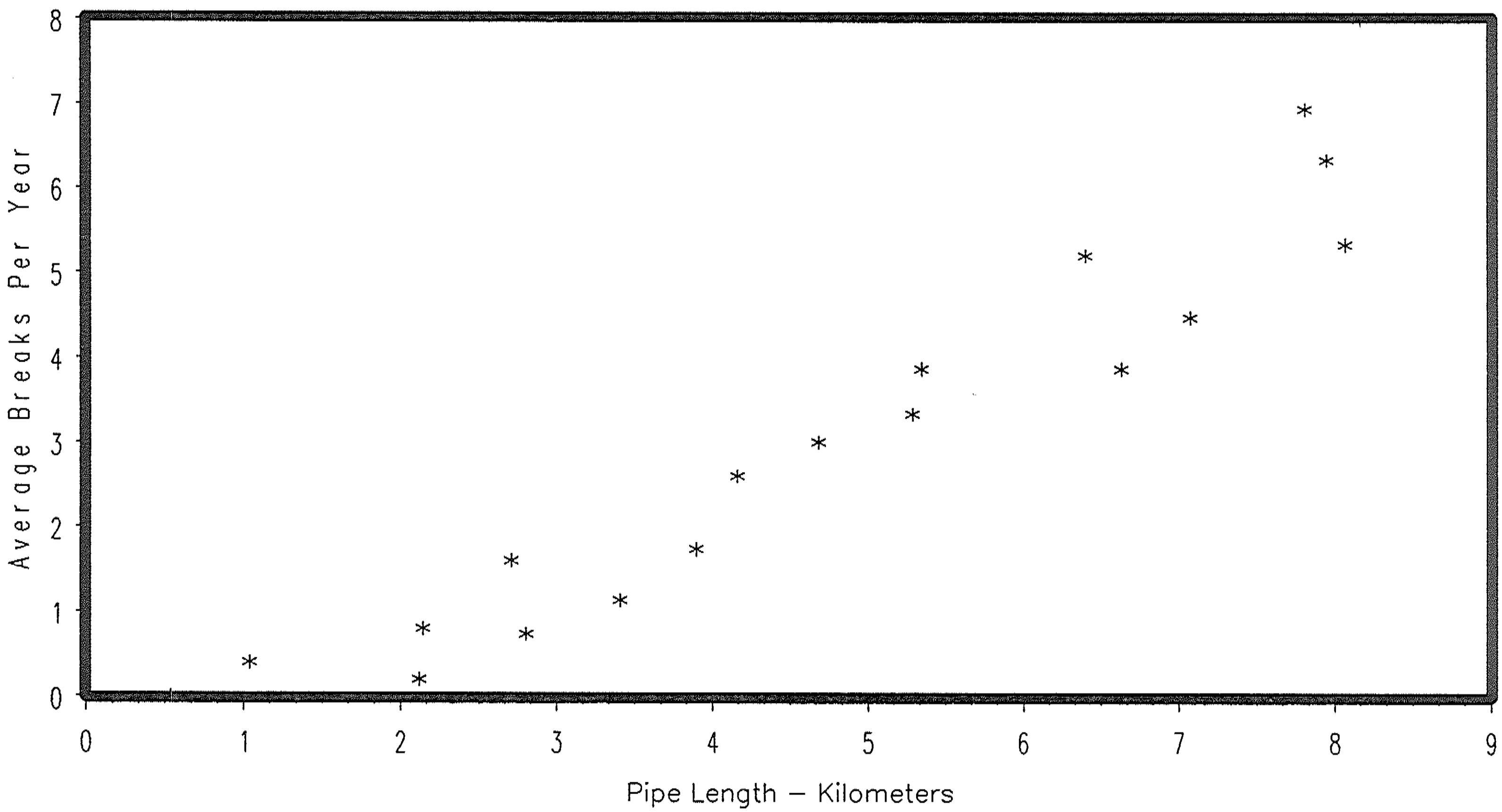


FIGURE B.13 Average number of breaks per km: asbestos cement, 200mm 31+ years

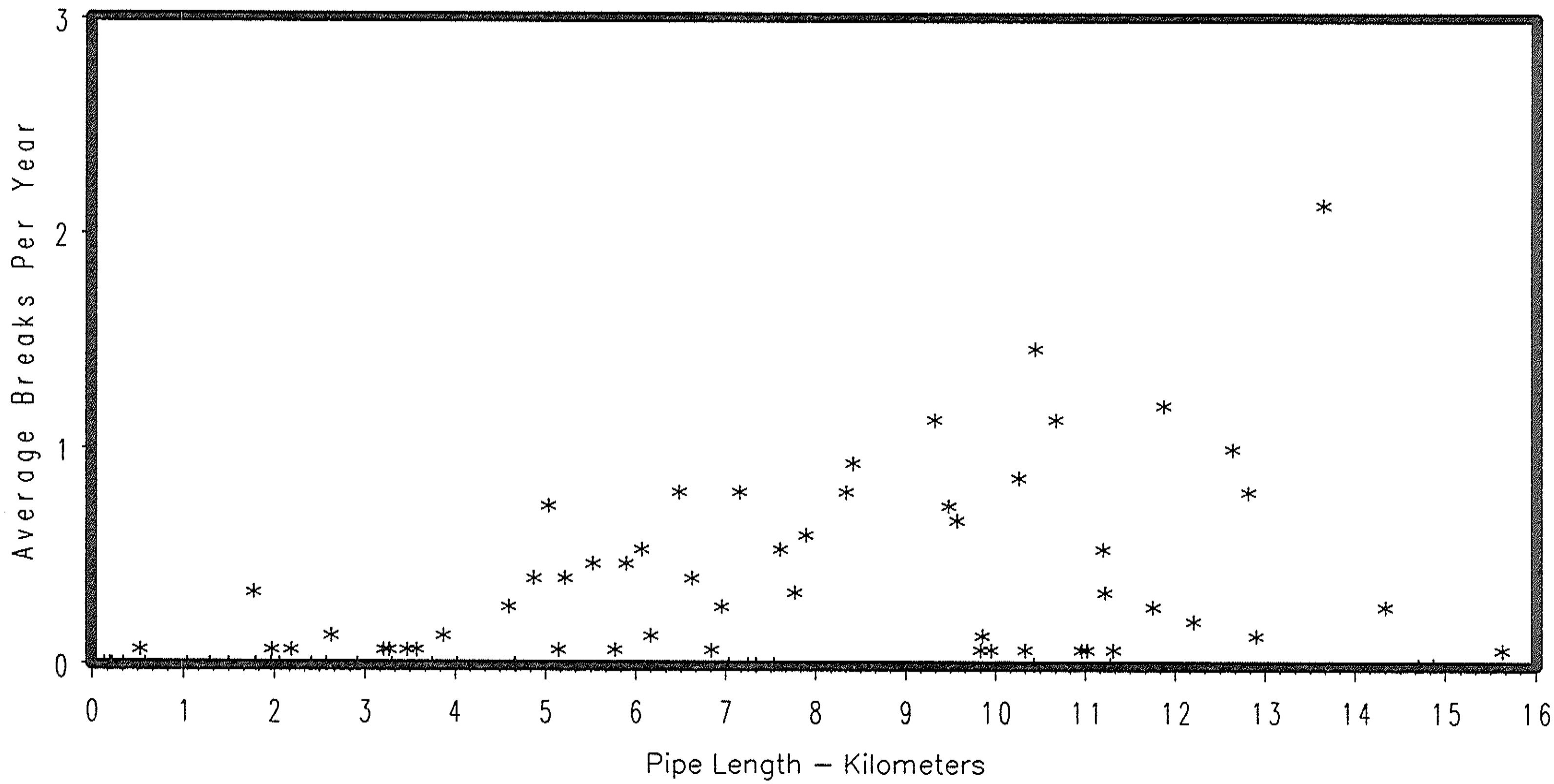


FIGURE B.14 Average number of breaks per km: asbestos cement, 250mm 0+ years
250mm 0+ years

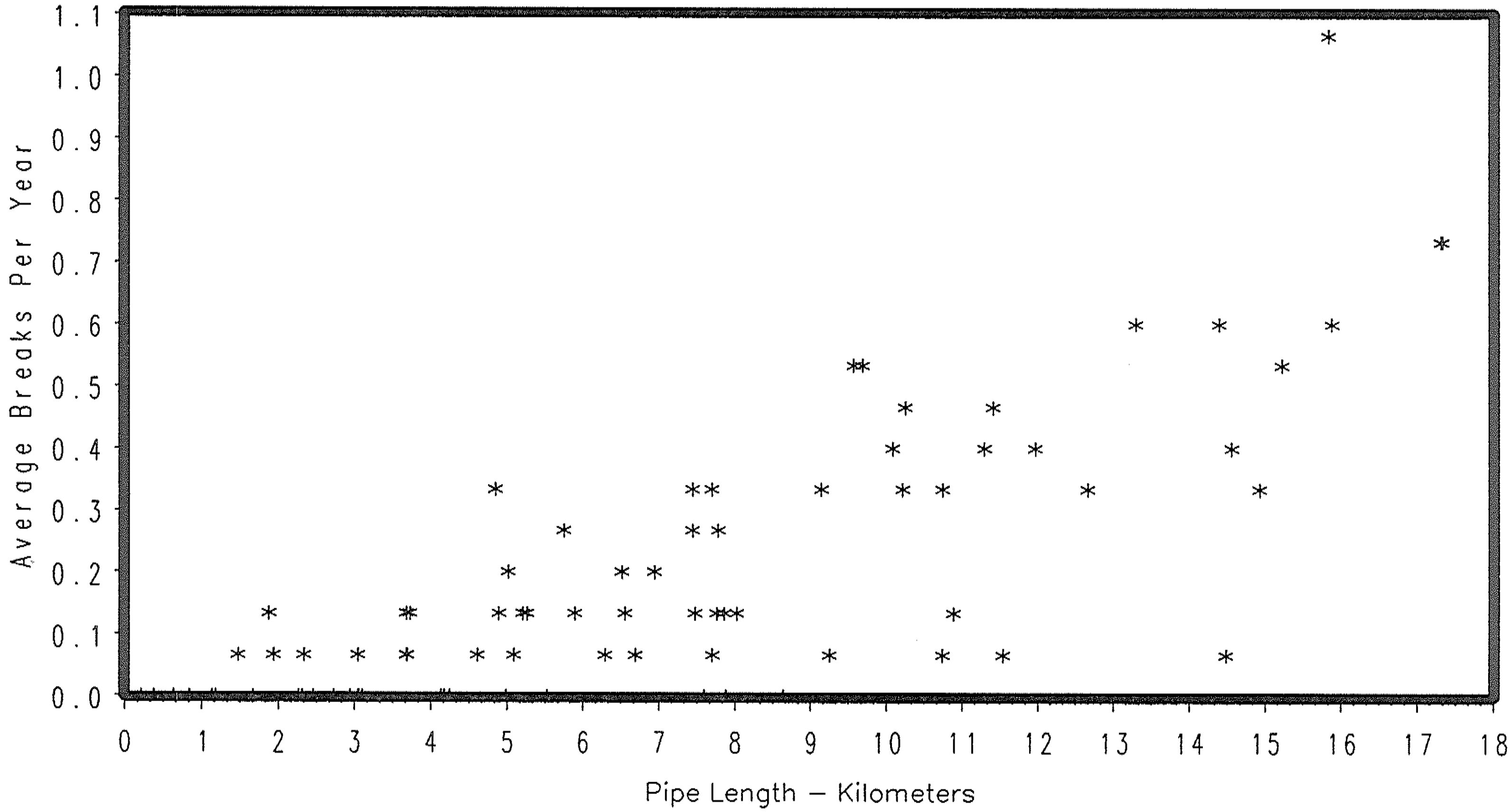


FIGURE B.15 Average number of breaks per km: asbestos cement, 300mm 0+ years

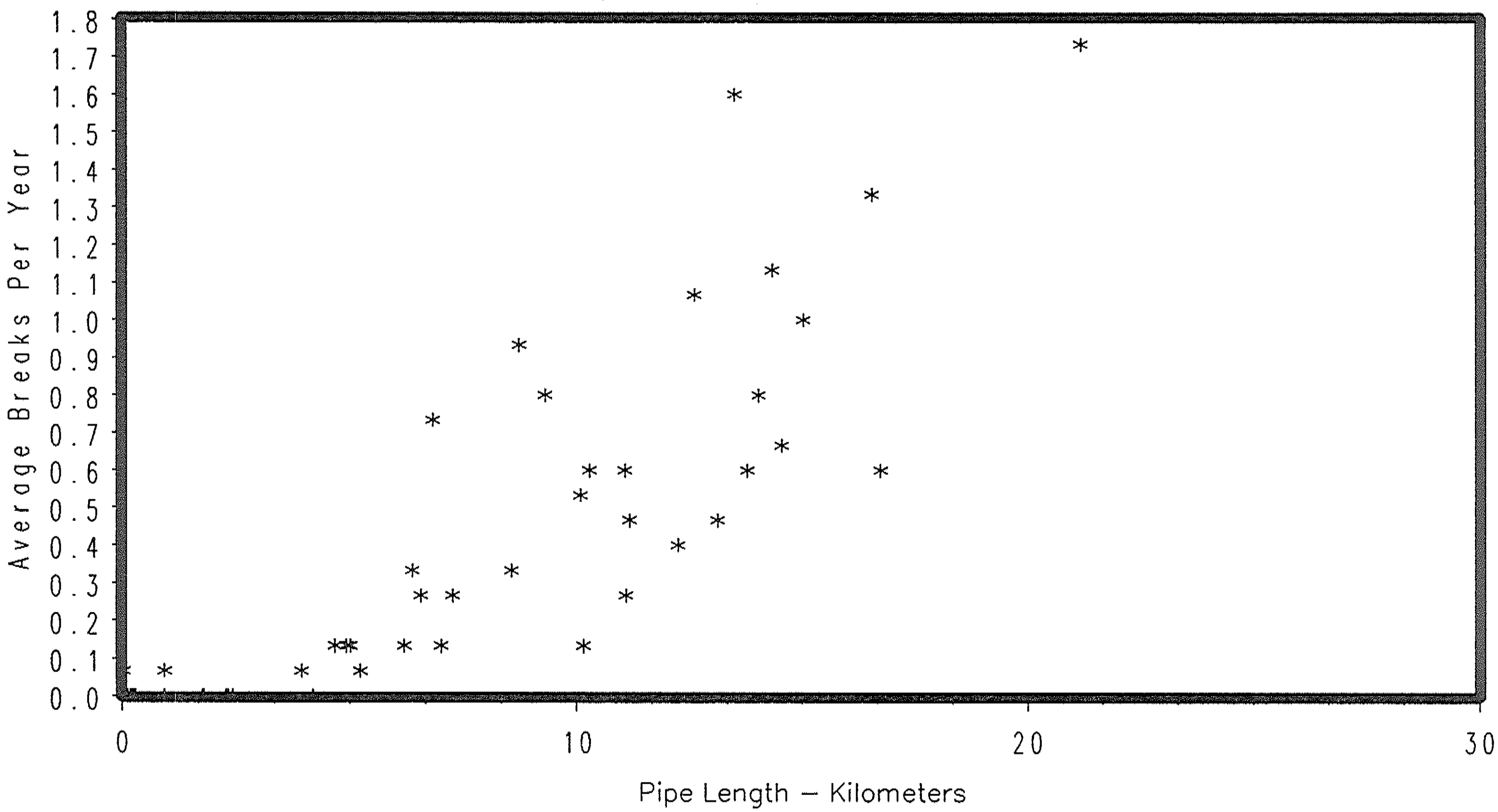


FIGURE B.16 Average number of breaks per km: ductile steel pipe, 150mm 0+ years

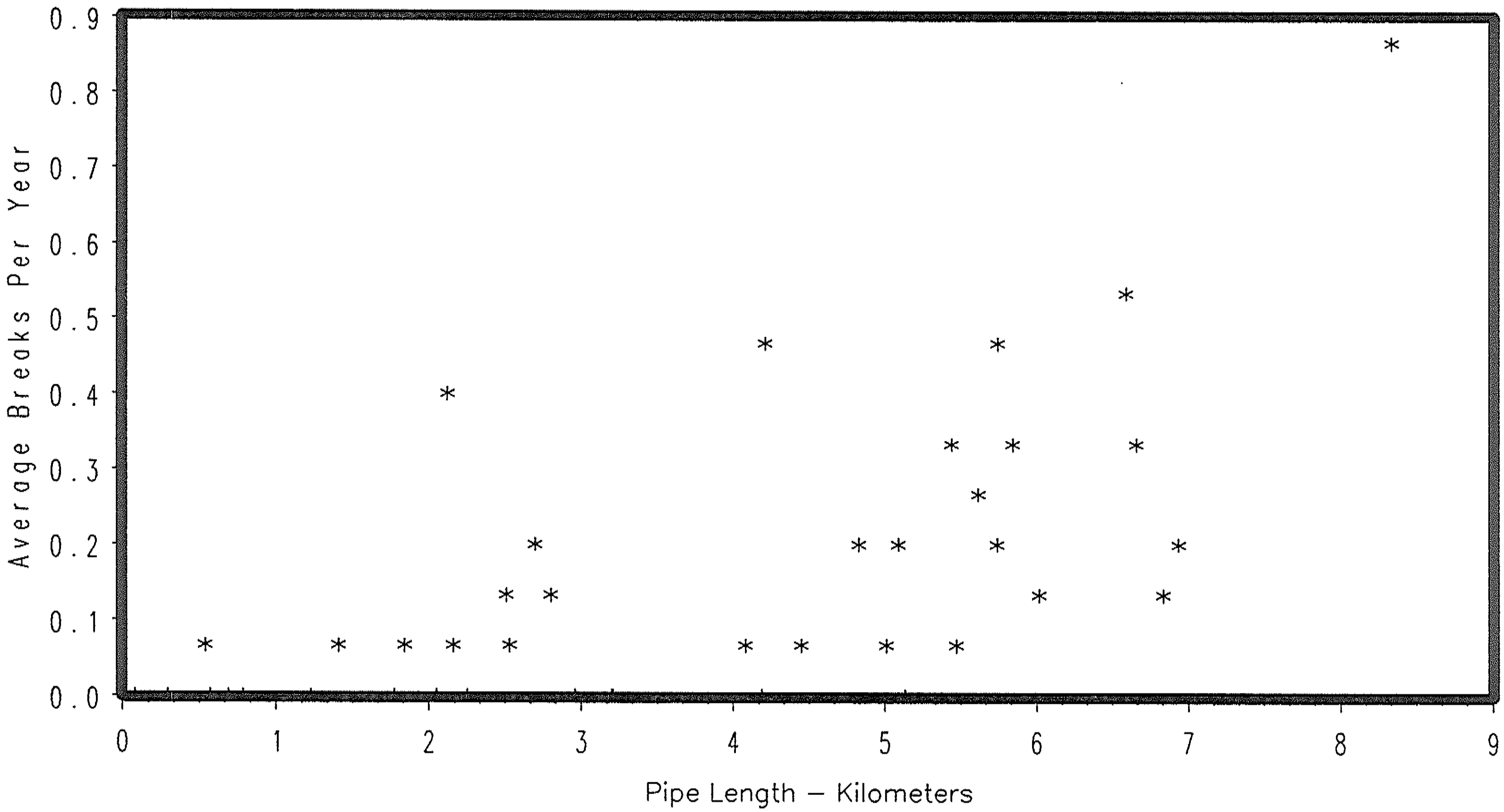


FIGURE B.17 Average number of breaks per yr: ductile steel pipe, 200mm 0+ years

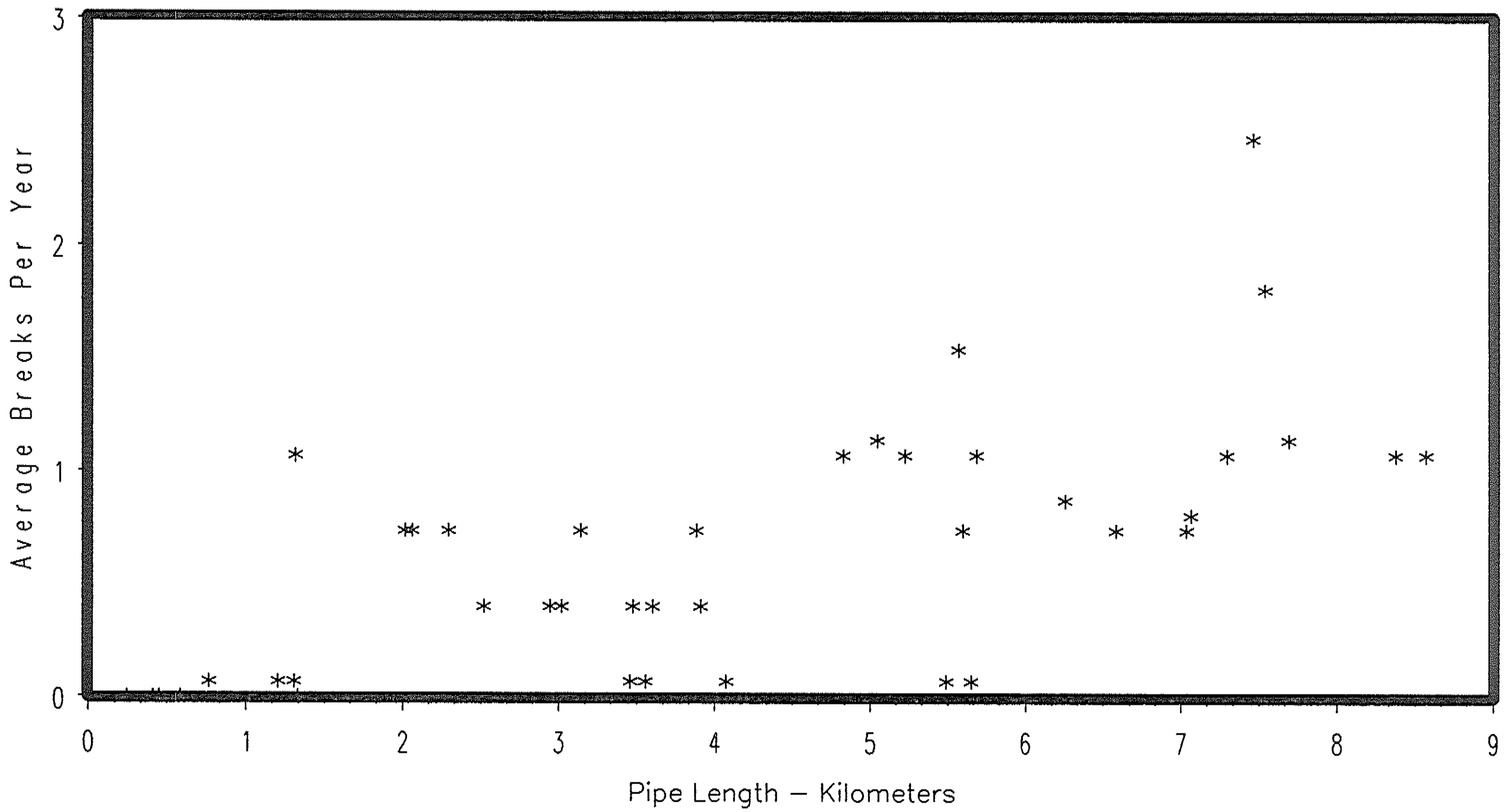


FIGURE B.18 Average number of breaks per yr: ductile steel pipe, 250mm 0+ years

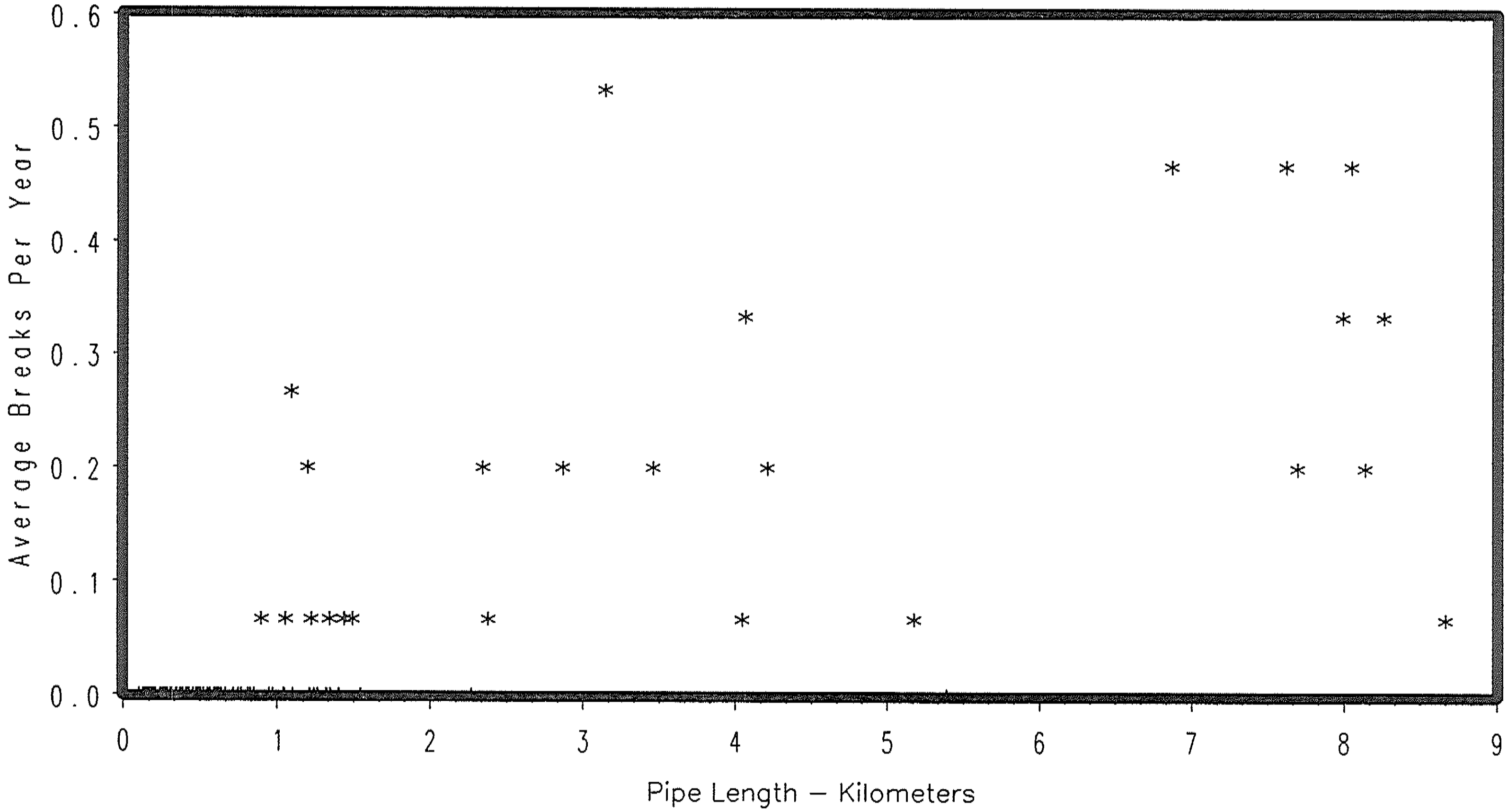


FIGURE B.19 Average number of breaks per yr: PVC pipe, 150mm 0+ years

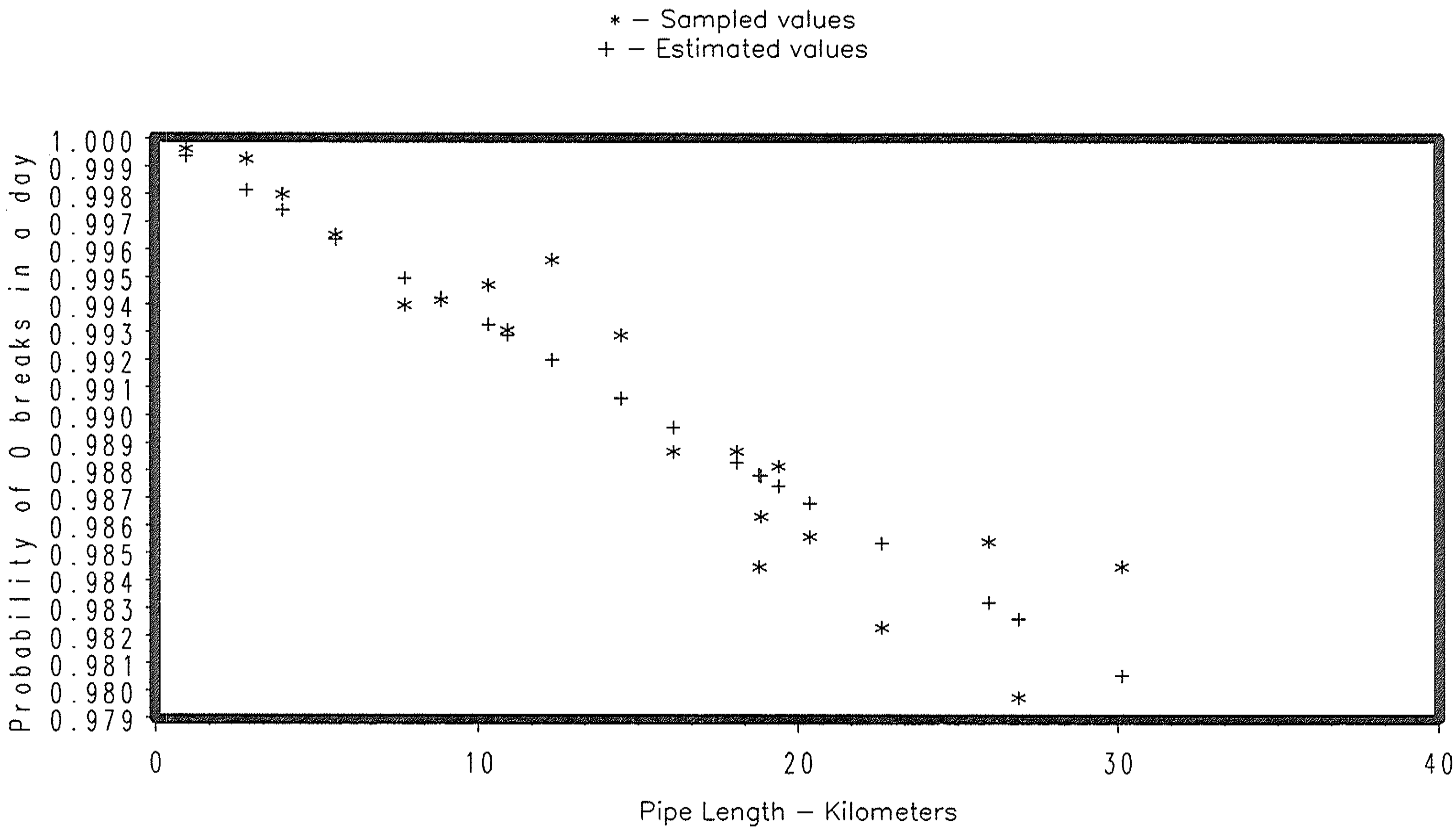


FIGURE B.20 Poisson estimated probability for m=0 breaks: cast iron pipe, 150mm 20–30 years

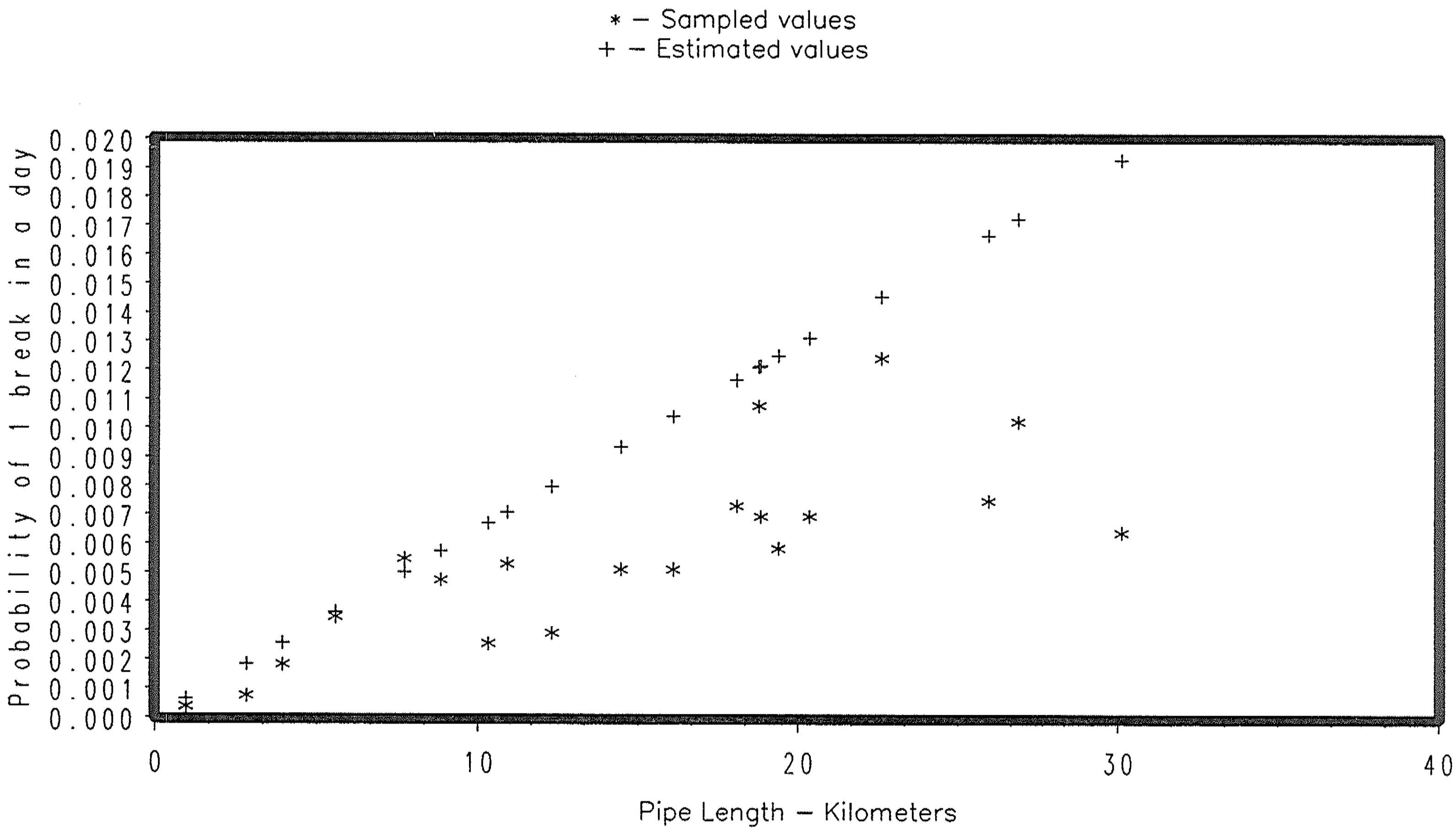


FIGURE B.21 Poisson estimated probability for m=1 break: cast iron pipe, 150mm 20–30 years

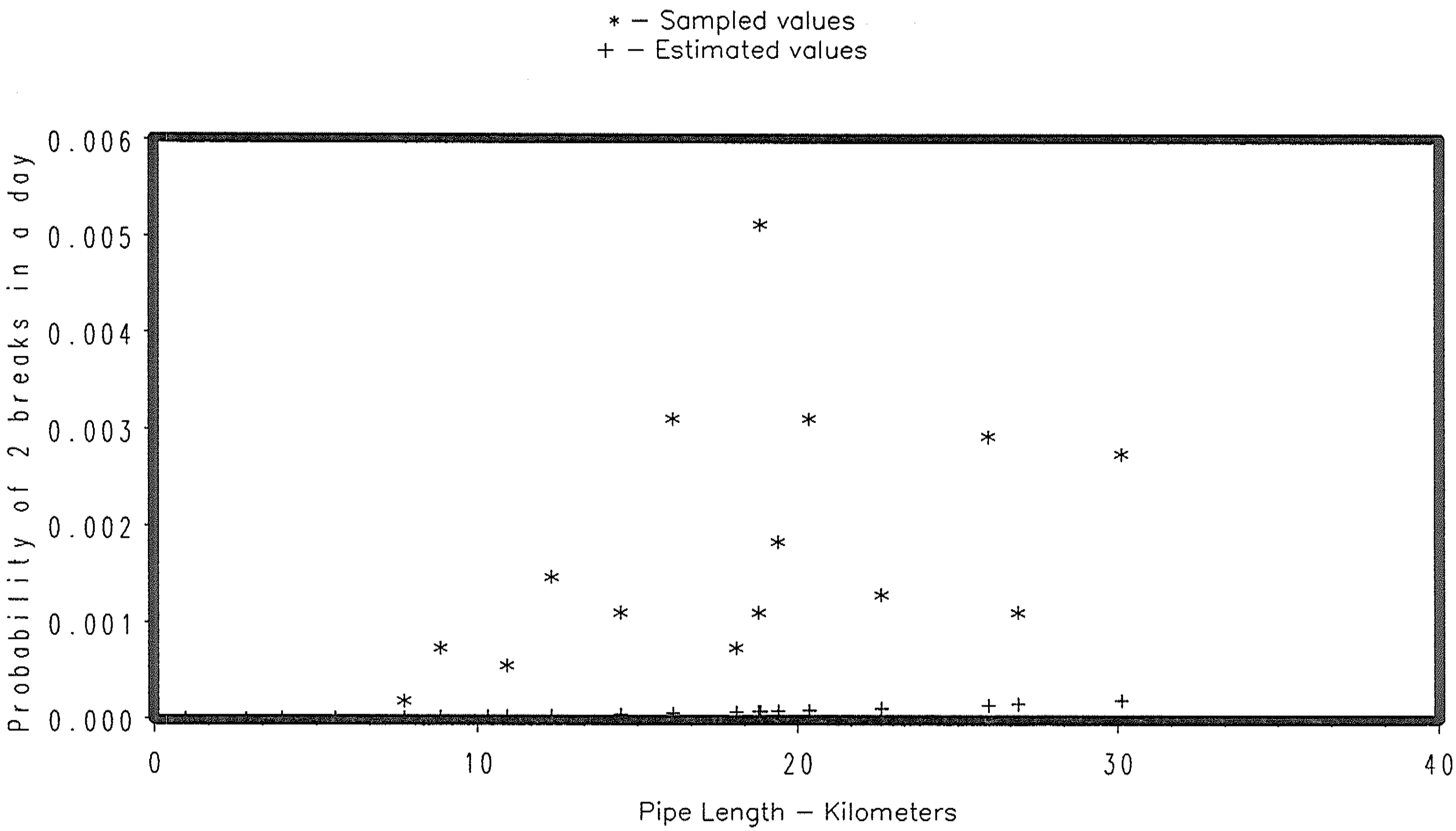


FIGURE B.22 Poisson estimated probability for m=2 breaks: cast iron pipe, 150mm 20–30 years

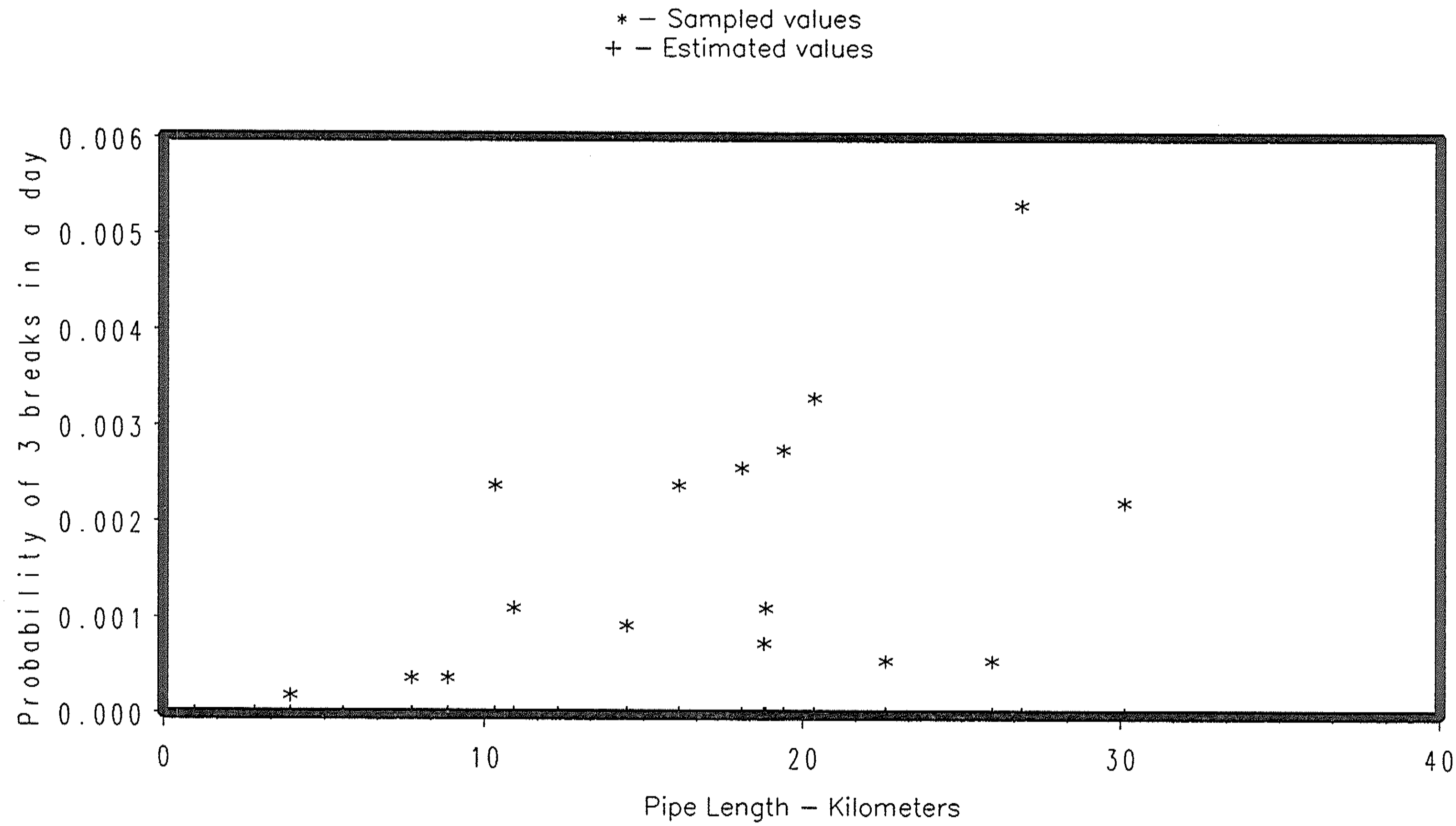


FIGURE B.23 Poisson estimated probability for m=3 breaks: cast iron pipe, 150mm 20–30 years

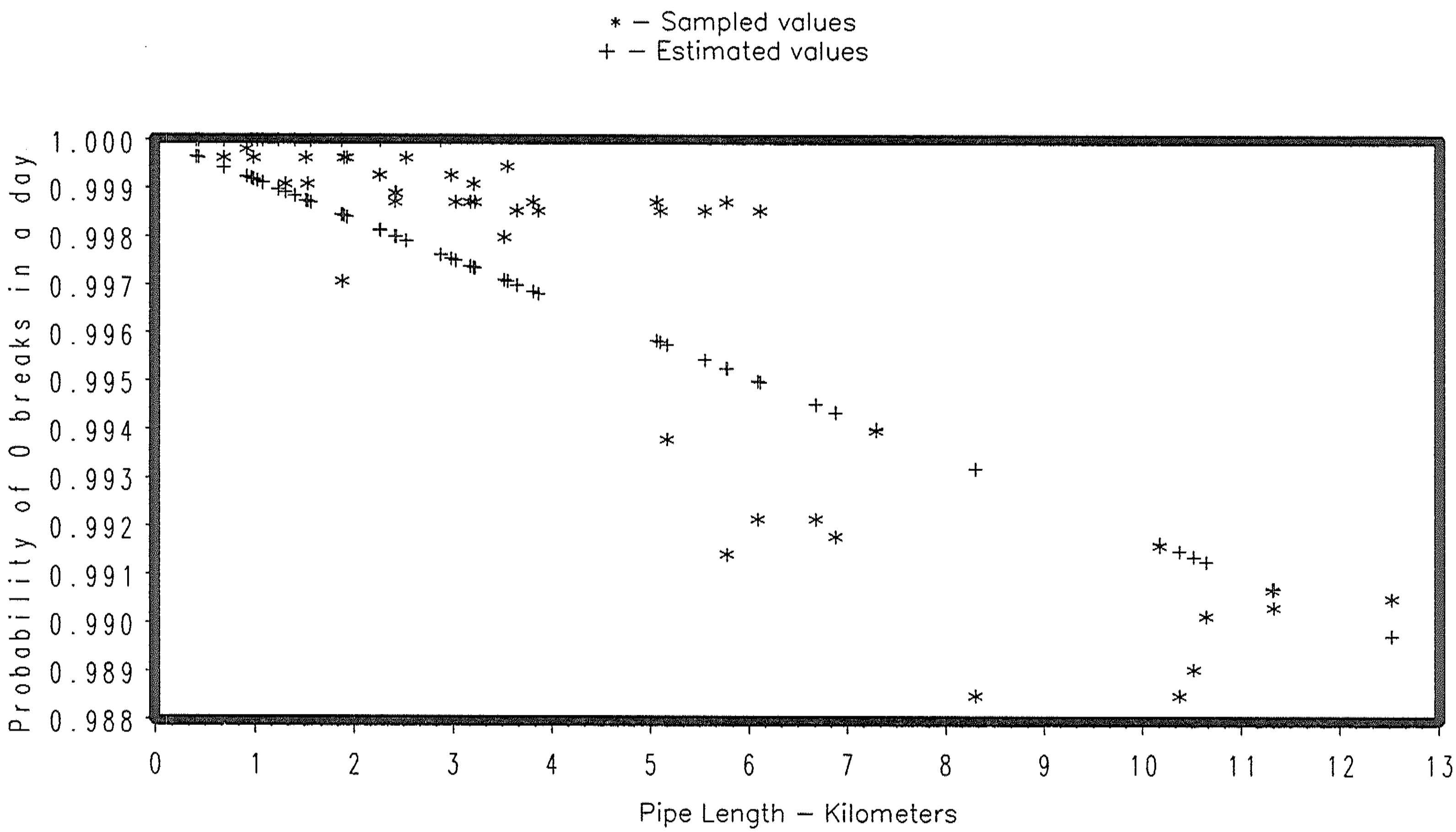


FIGURE B.24 Poisson estimated probability for m=0 breaks: cast iron pipe, 150mm 31+ years

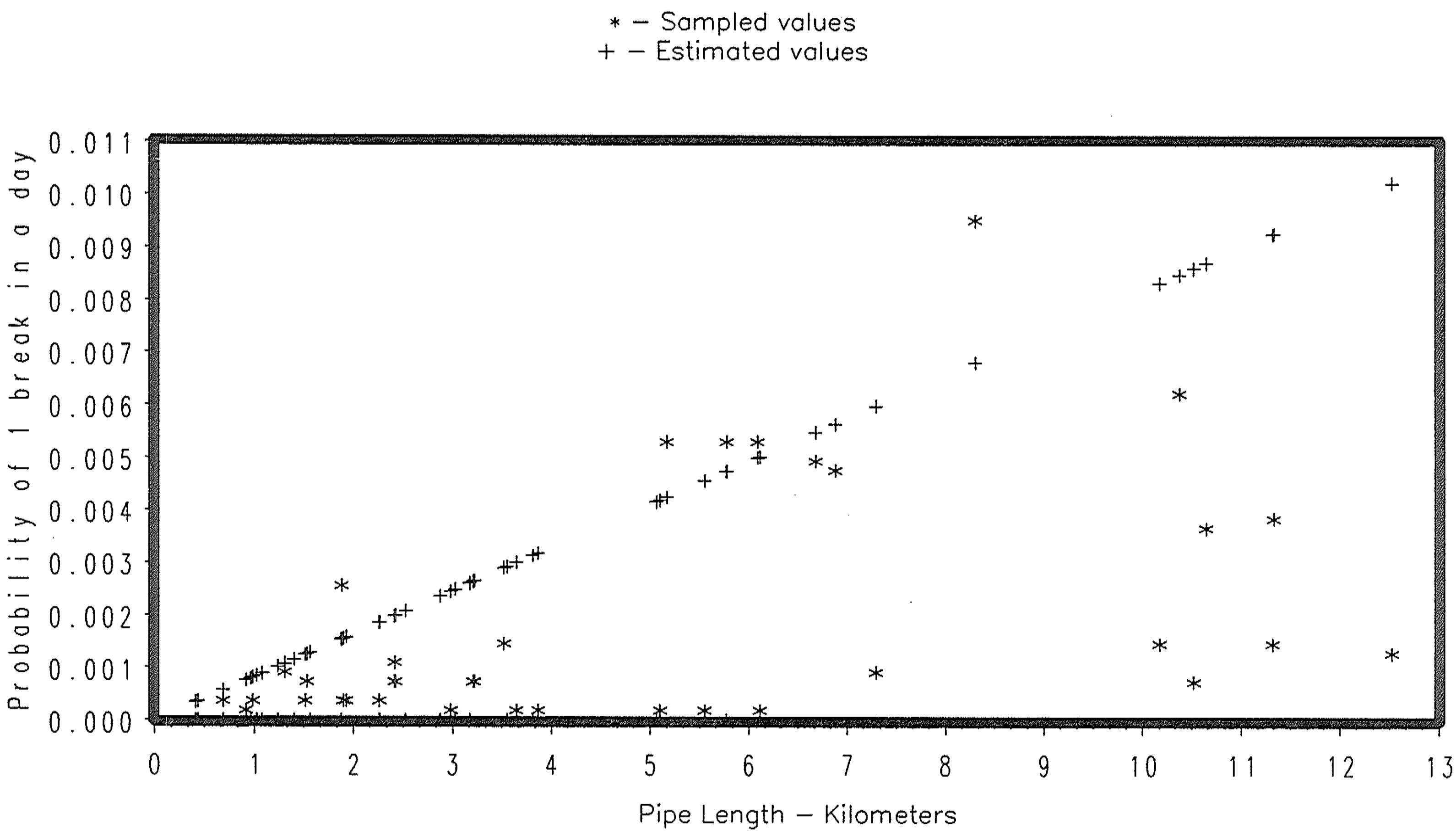


FIGURE B.25 Poisson estimated probability for m=1 break: cast iron pipe, 150mm 31+ years

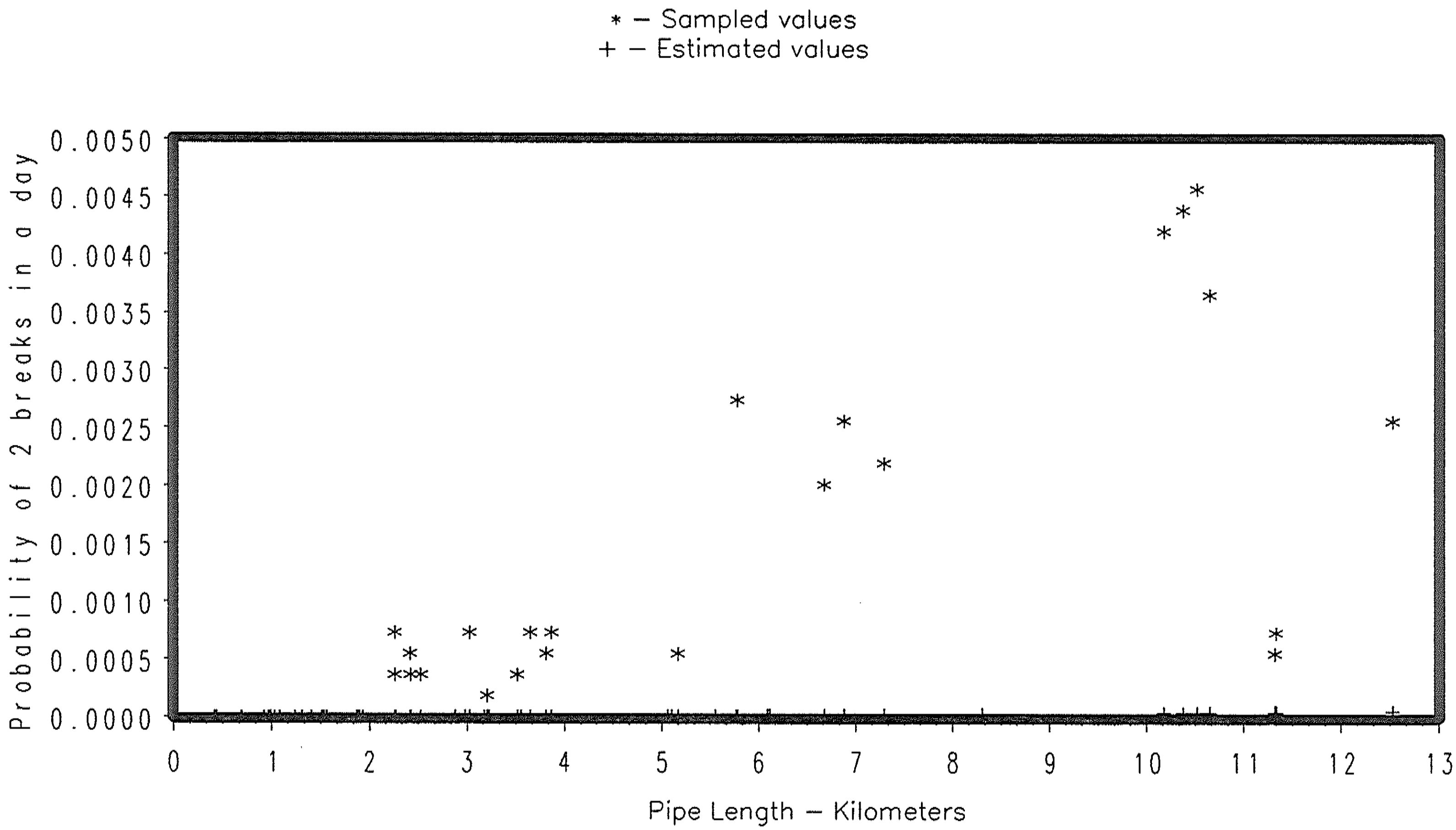


FIGURE B.26 Poisson estimated probability for m=2 breaks: cast iron pipe, 150mm 31+ years

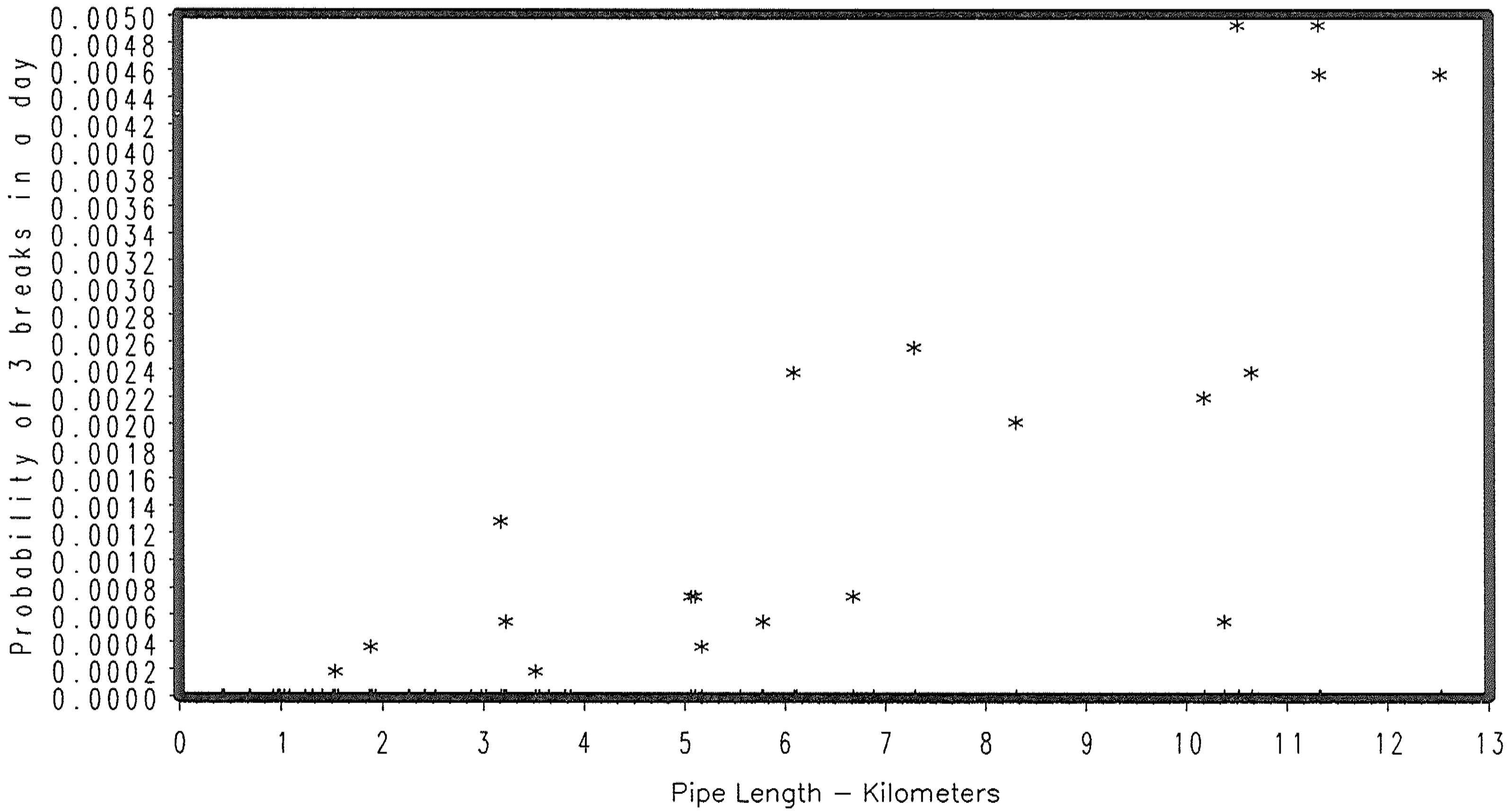


FIGURE B.27 Poisson estimated probability for m=3 breaks: cast iron pipe, 150mm 31+ years

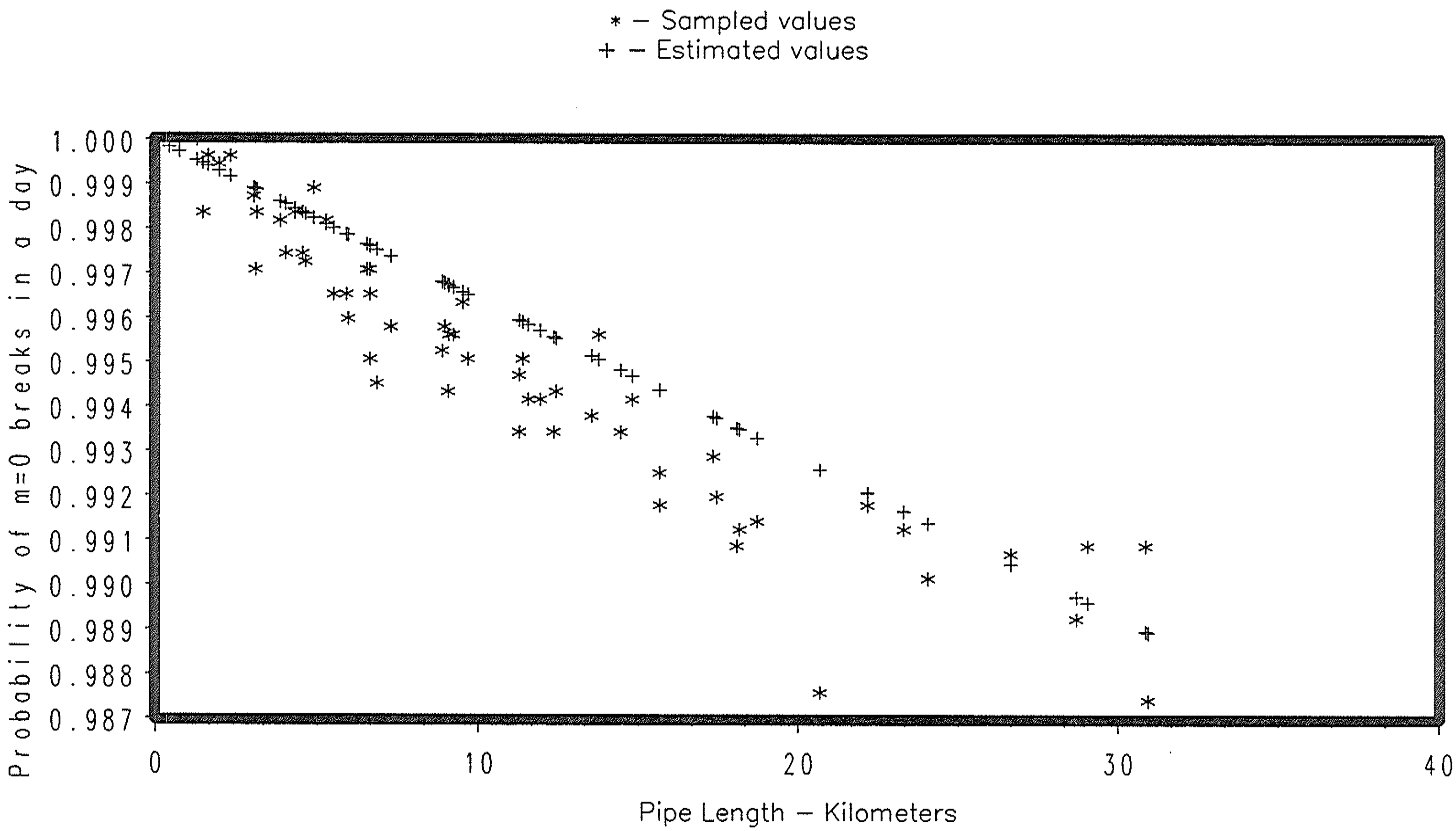


FIGURE B.28 Predicted versus sampled for m=0 breaks: cast iron pipe, 150mm 0–19 years

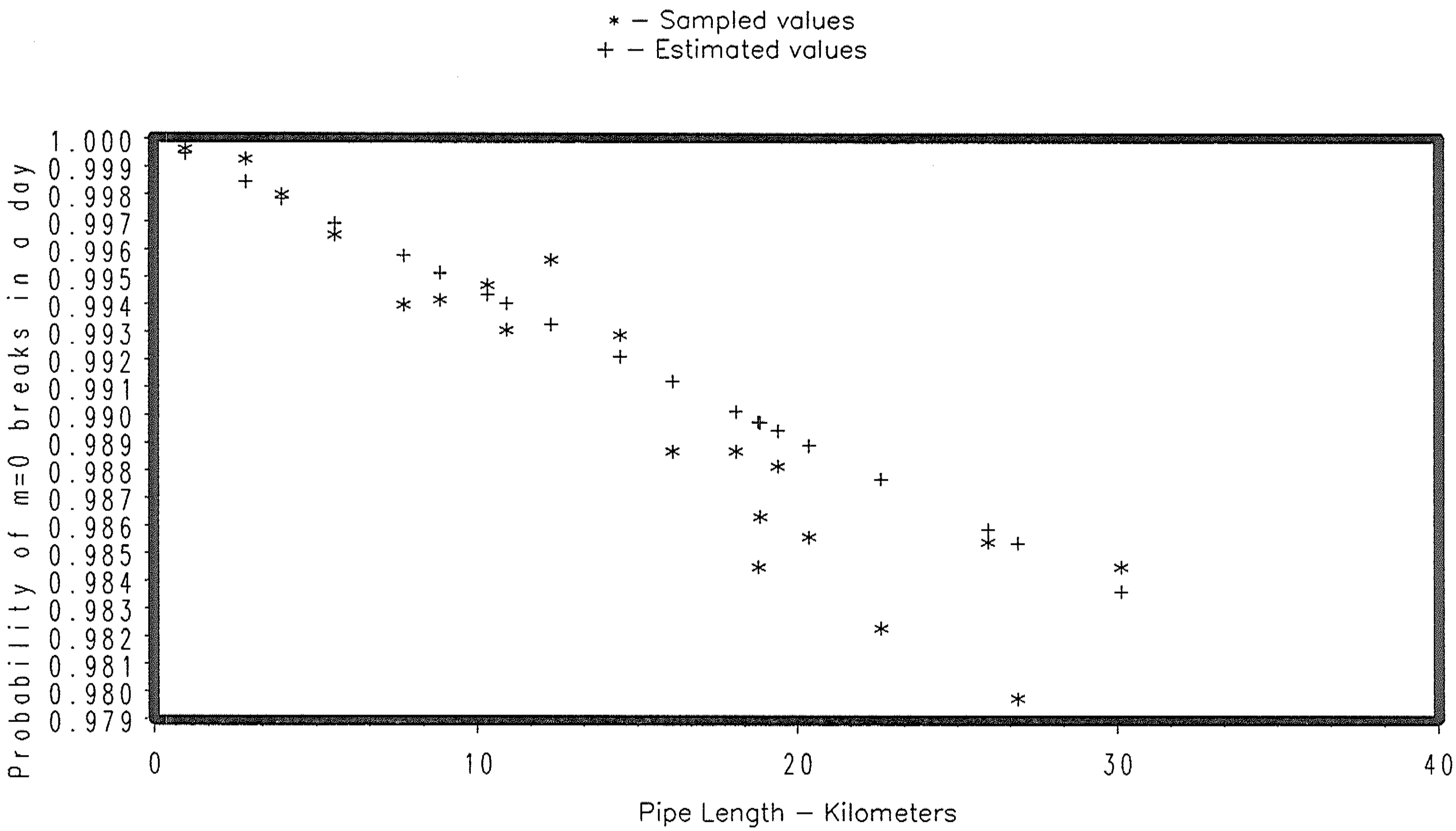


FIGURE B.29 Predicted versus sampled for m=0 breaks: cast iron pipe, 150mm 20–30 years

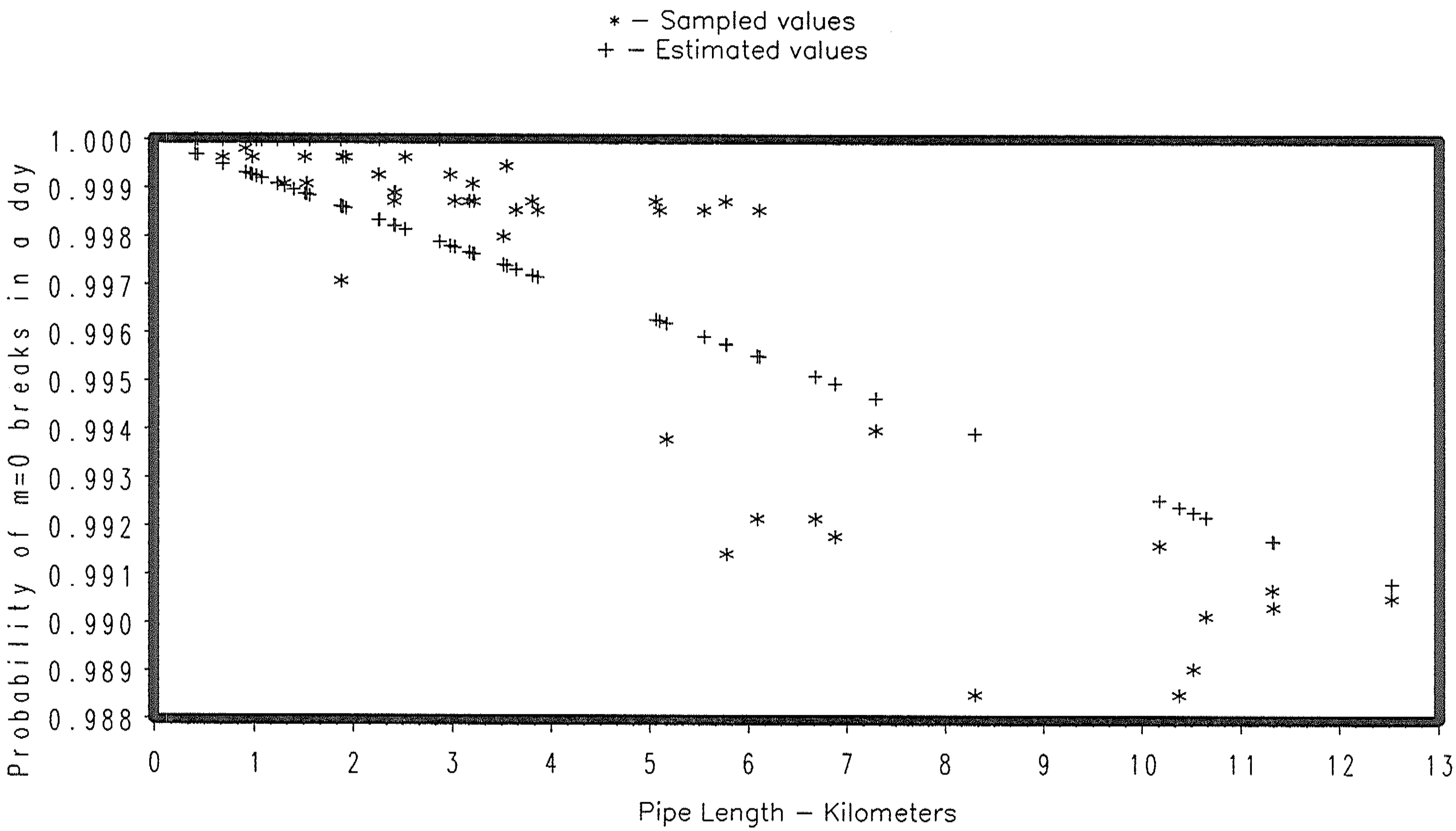


FIGURE B.30 Predicted versus sampled for m=0 break: cast iron pipe, 150mm 31+ years

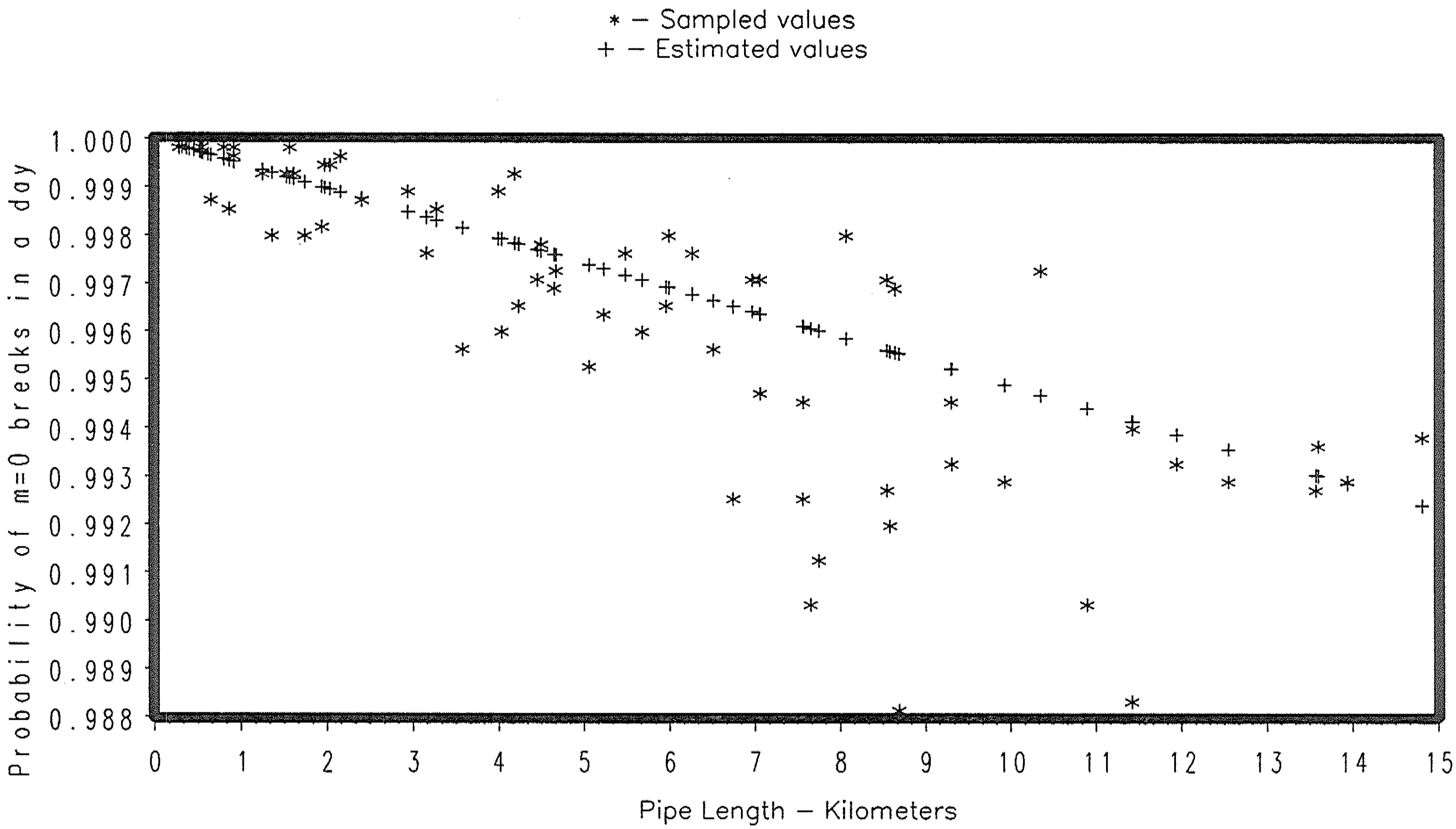


FIGURE B.31 Predicted versus sampled for m=0 break: cast iron pipe, 200mm 0–30 years

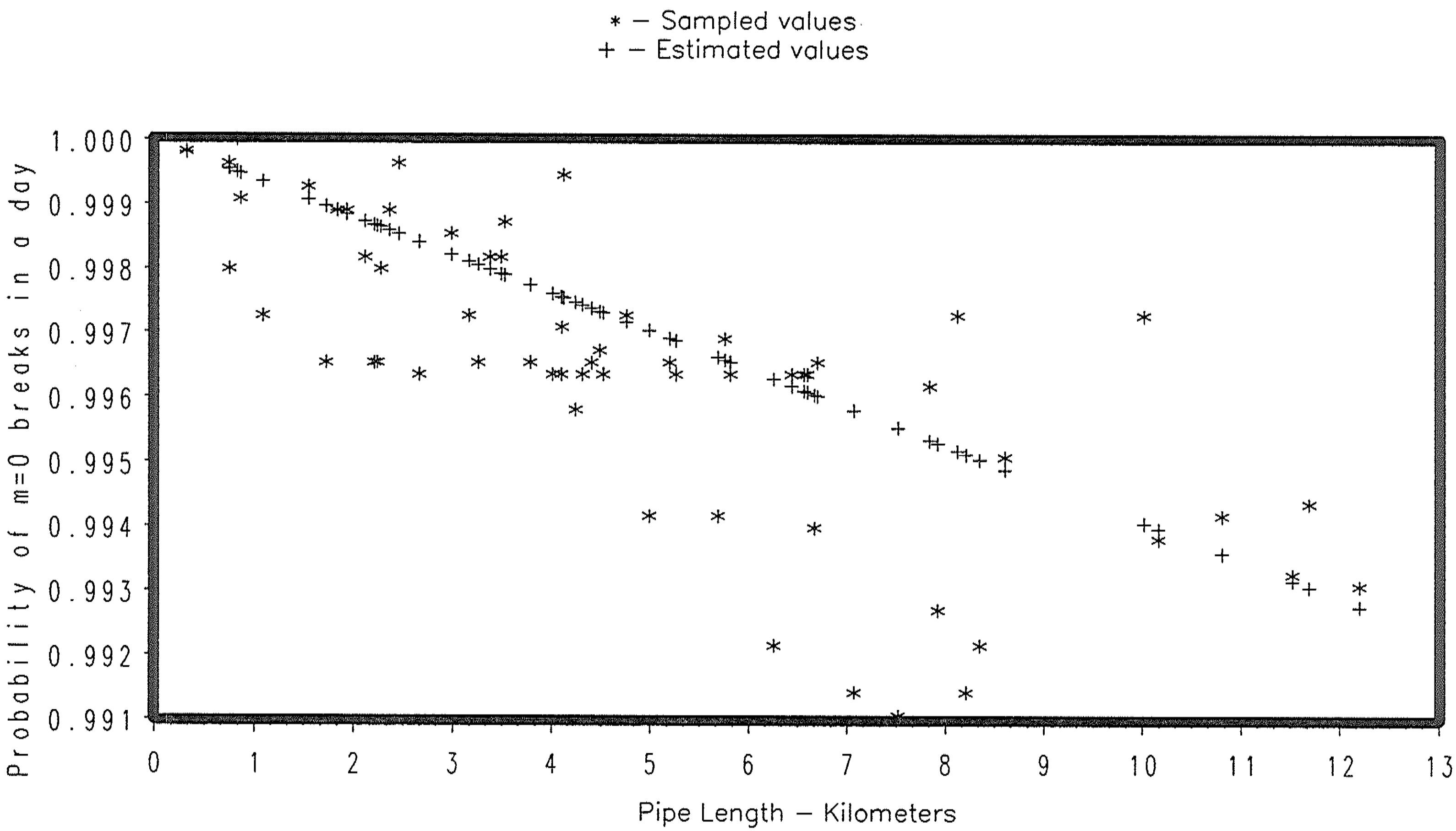


FIGURE B.32 Predicted versus sampled for m=0 break: cast iron pipe, 200mm 31+ years

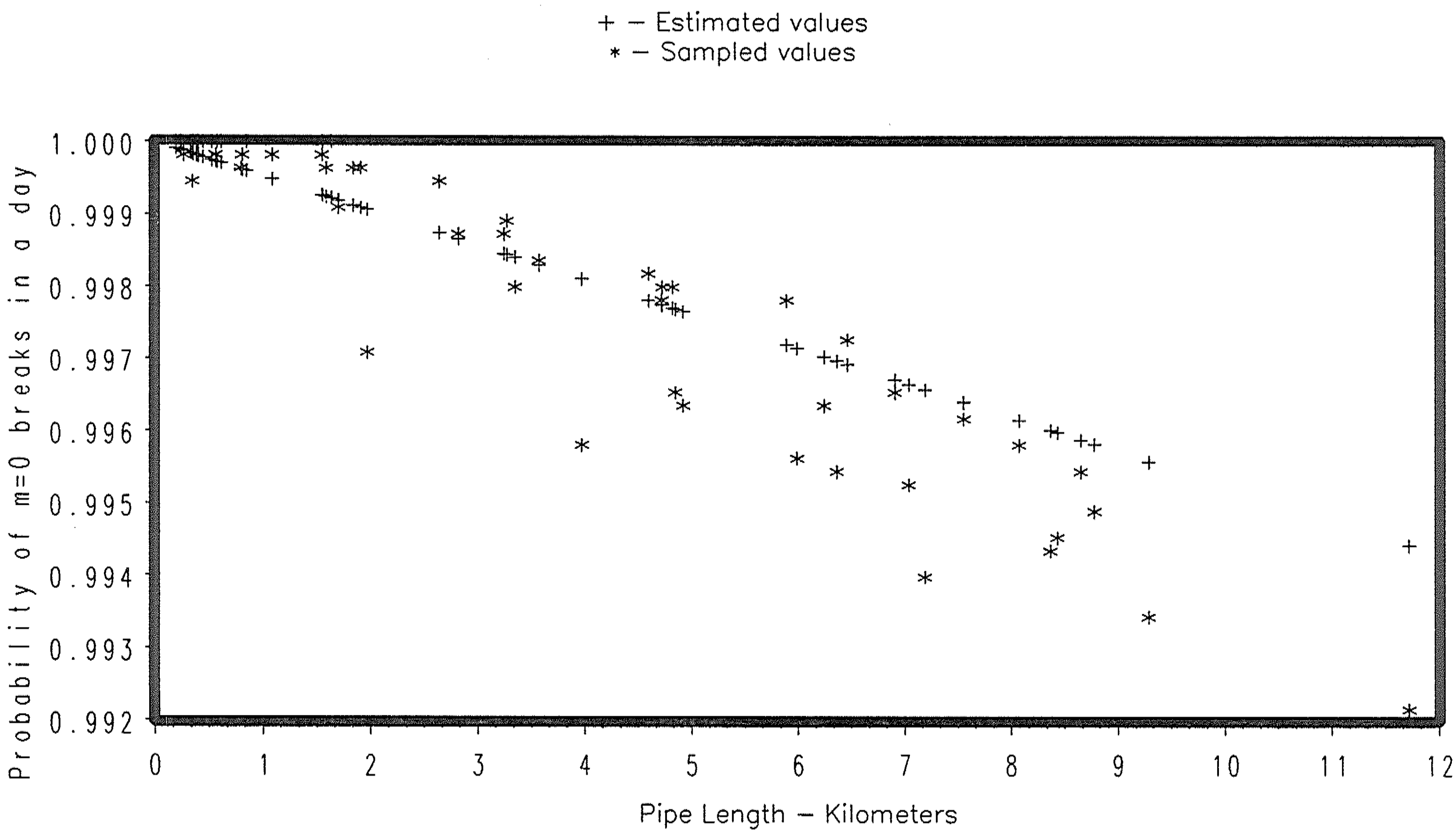


FIGURE B.33 Predicted versus sampled for m=0 breaks: cast iron pipe, 250mm 0–30 years

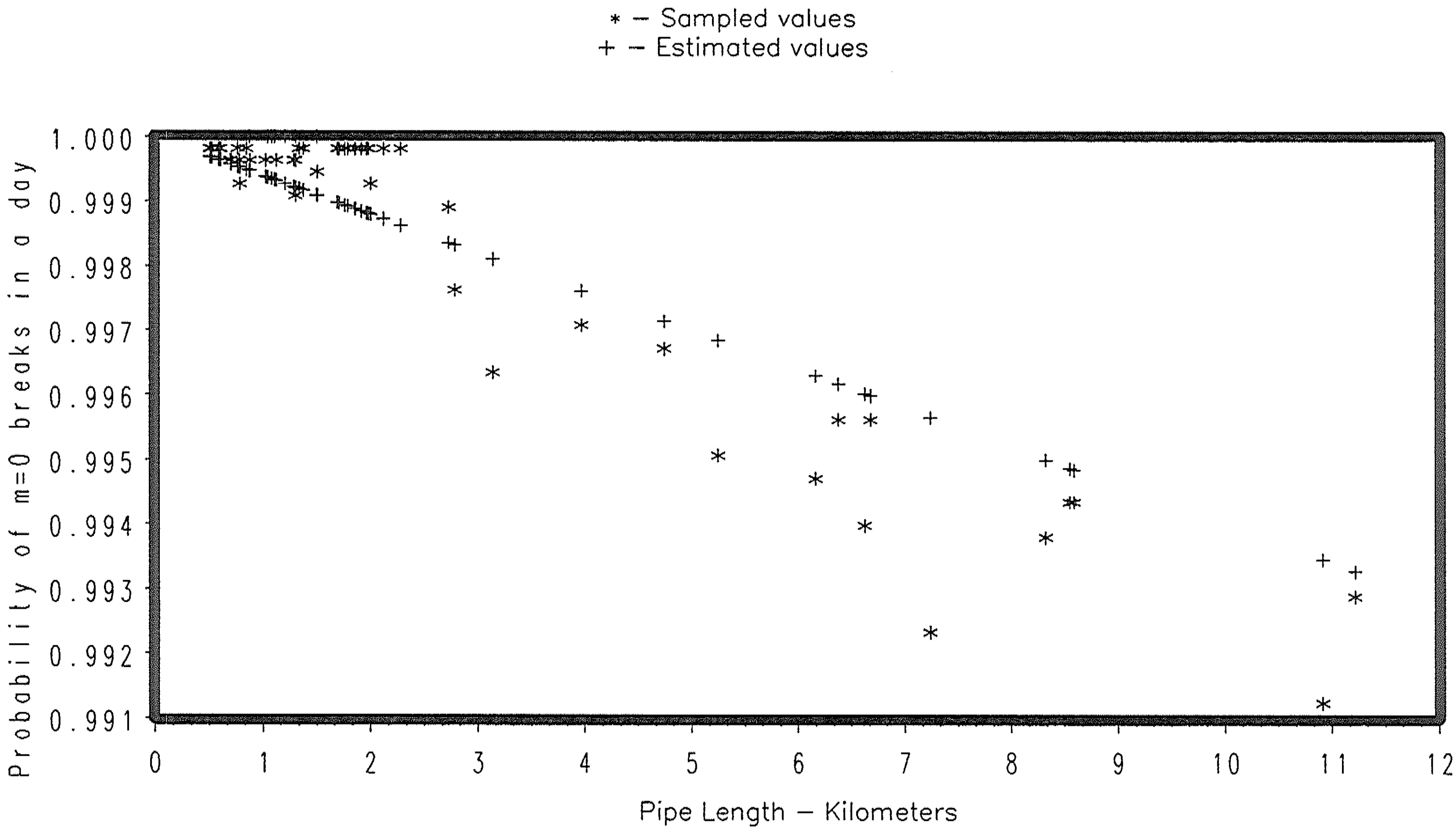


FIGURE B.34 Predicted versus sampled for m=0 breaks: cast iron pipe, 250mm 31+ years

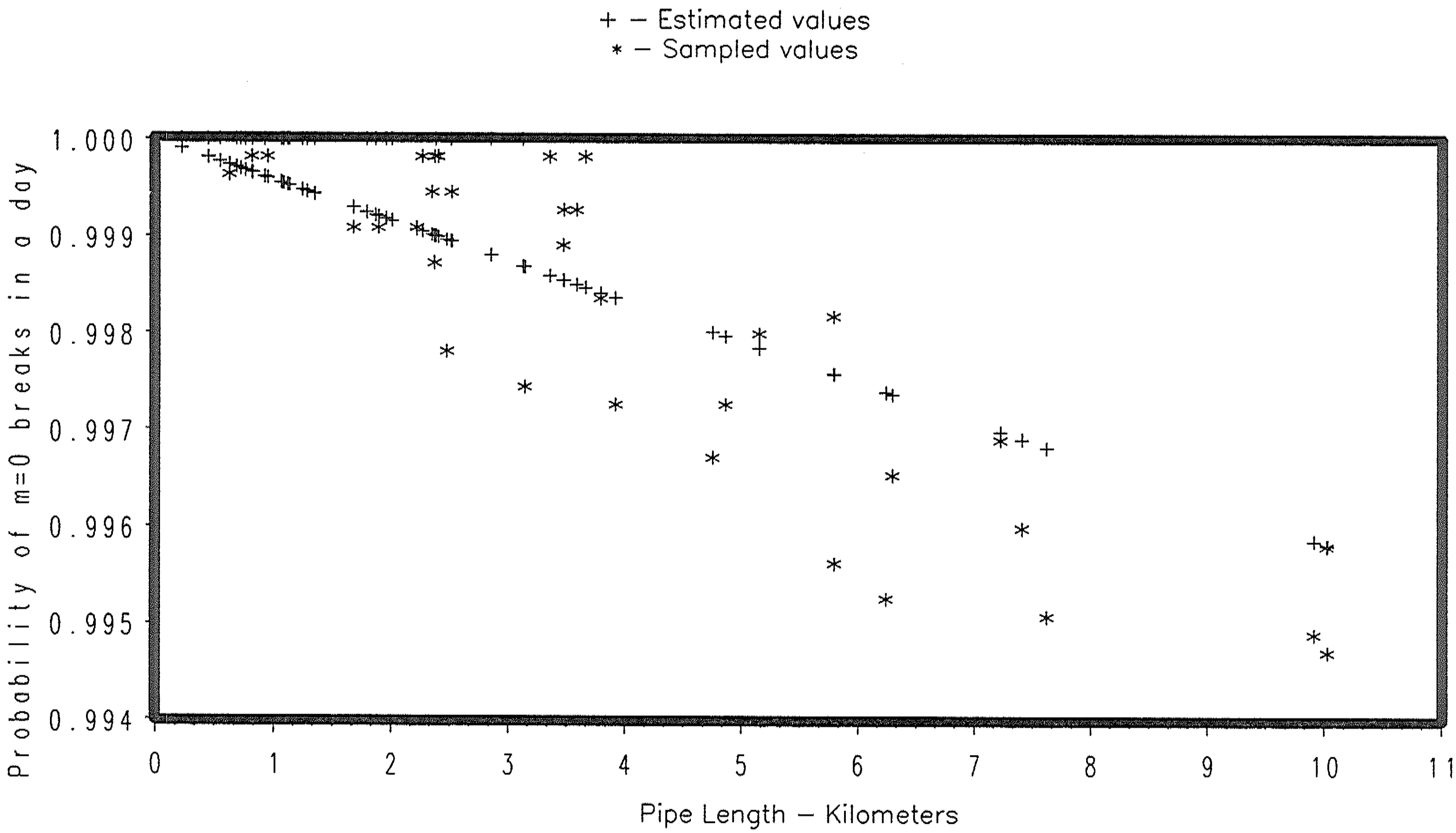


FIGURE B.35 Predicted versus sampled for m=0 breaks: cast iron pipe, 300mm 0–30 years

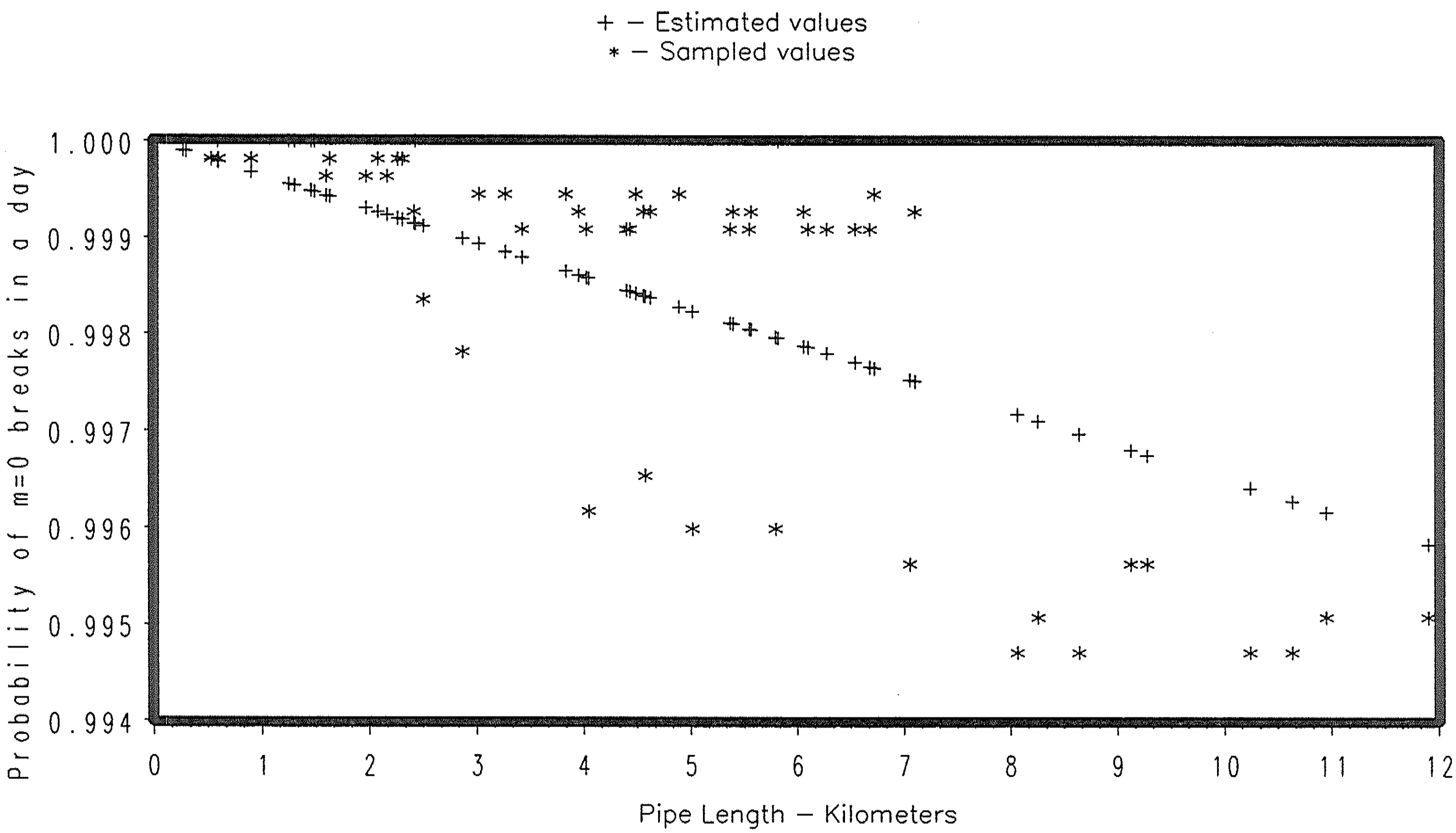


FIGURE B.36 Predicted versus sampled for m=0 breaks: cast iron pipe, 300mm 31+ years

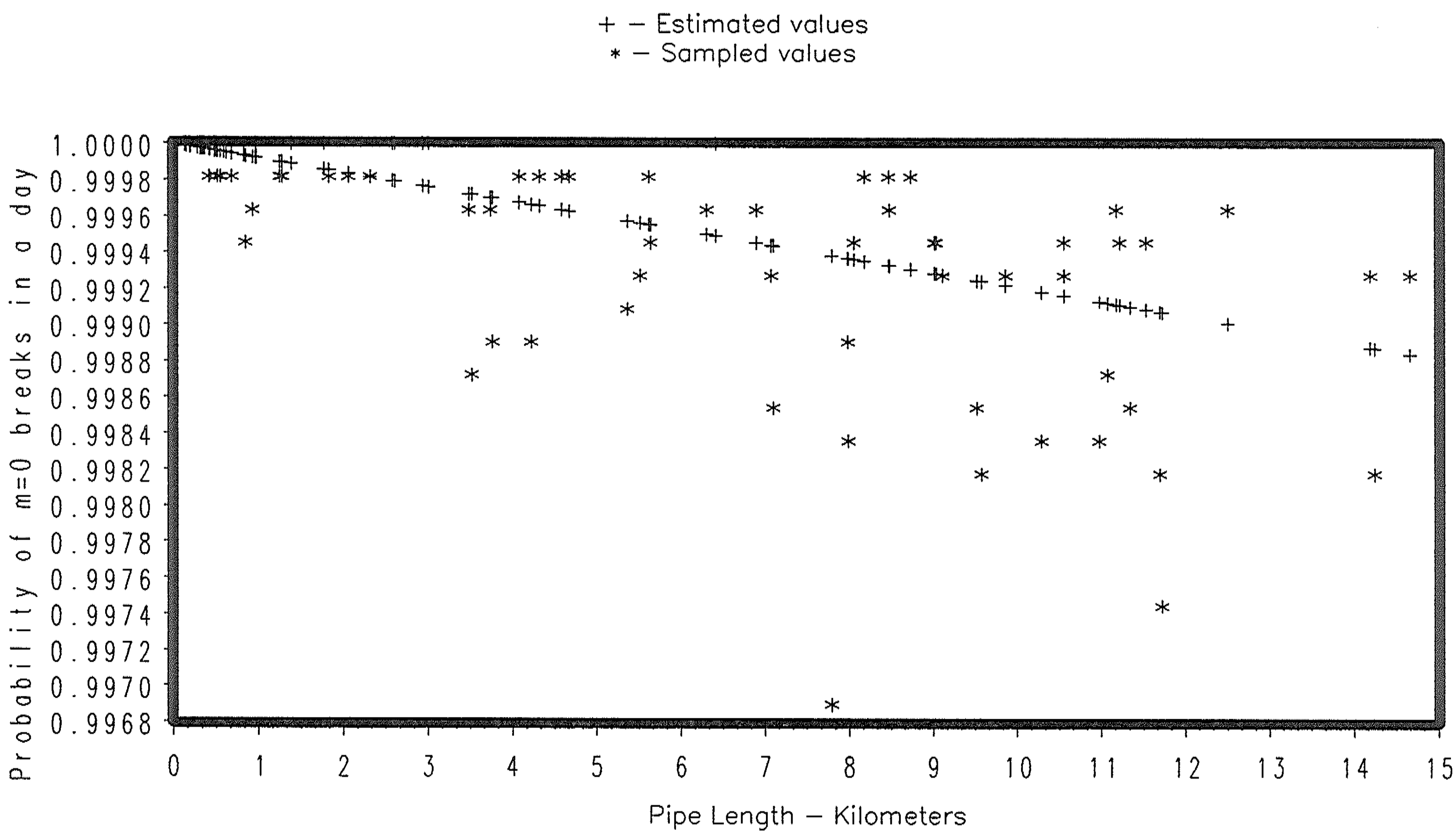


FIGURE B.37 Predicted versus sampled for m=0 breaks: asbestos cement, 150mm 0–30 years

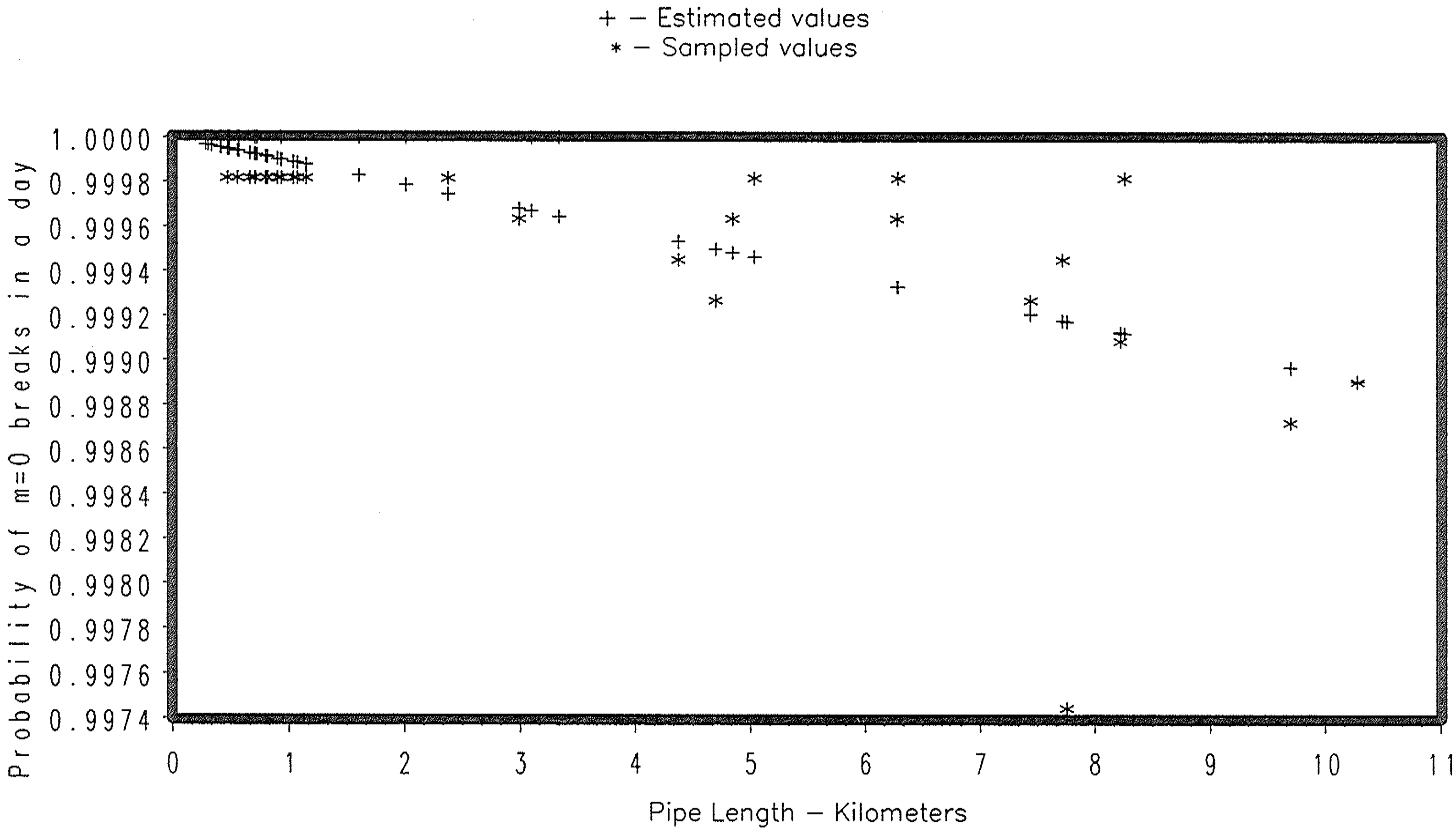


FIGURE B.38 Predicted versus sampled for m=0 breaks: asbestos cement, 150mm 30+ years

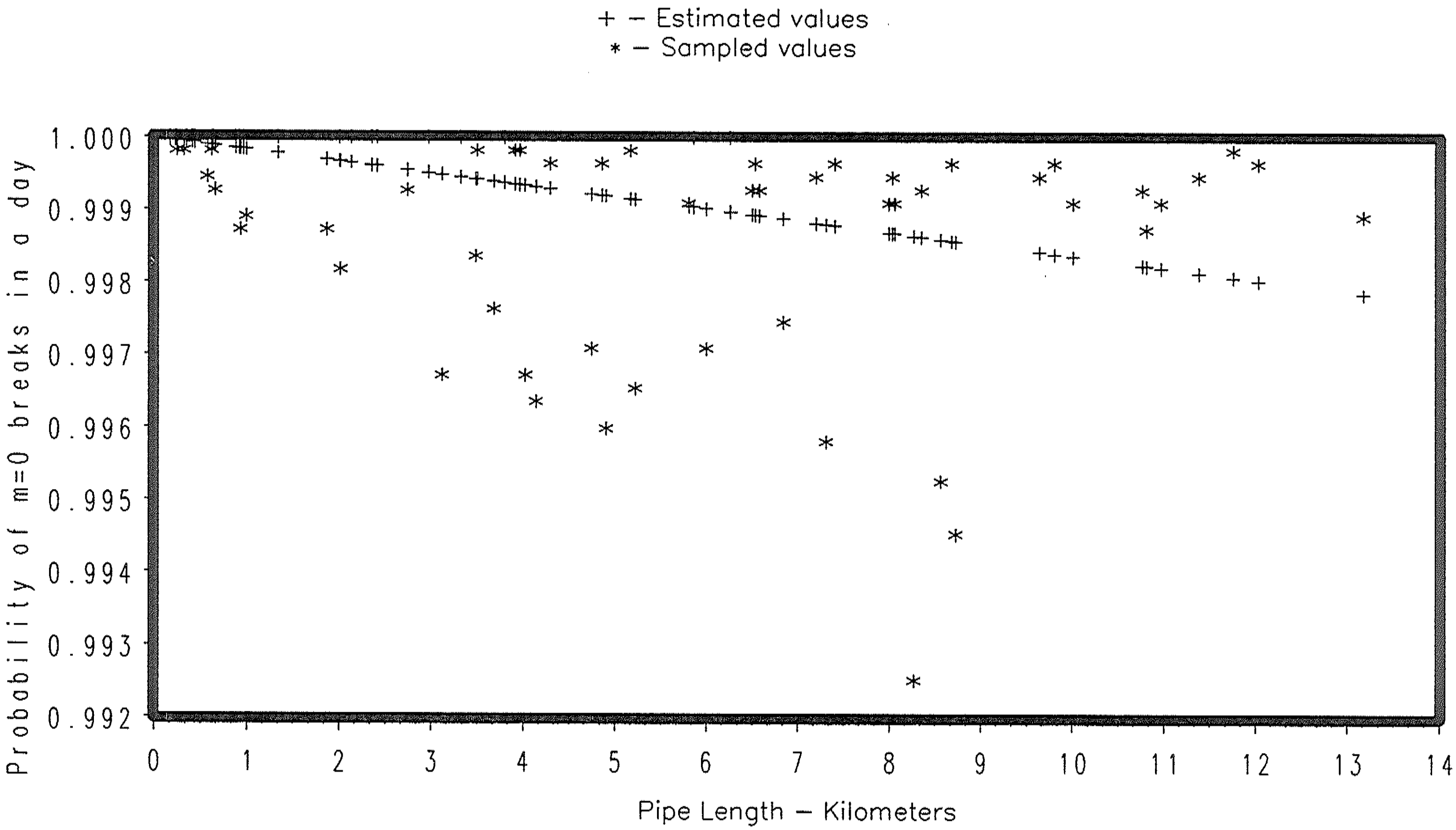


FIGURE B.39 Predicted versus sampled for m=0 breaks: asbestos cement, 200mm 0–30 years

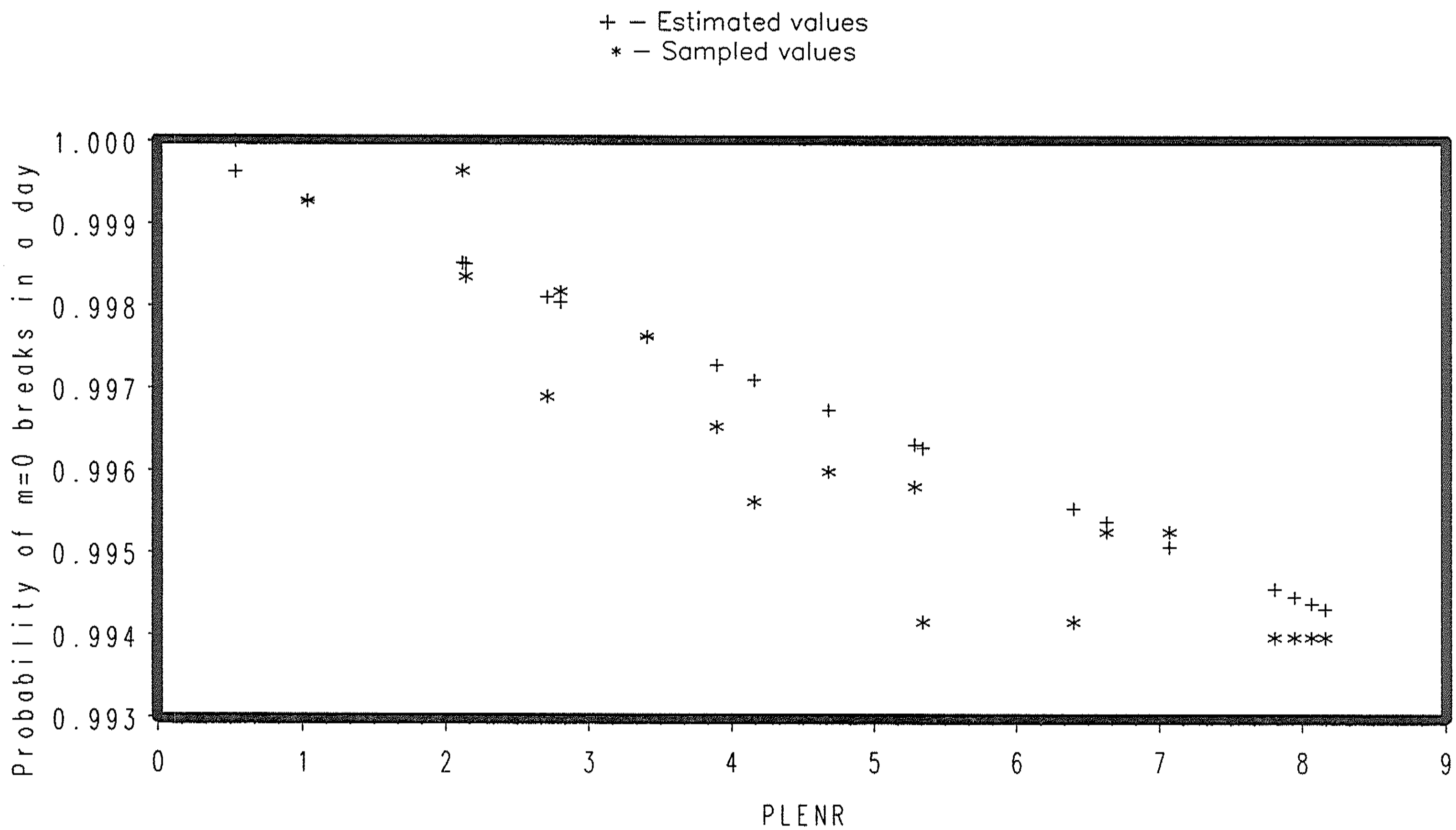


FIGURE B.40 Predicted versus sampled for m=0 breaks: asbestos cement, 200mm 31+ years

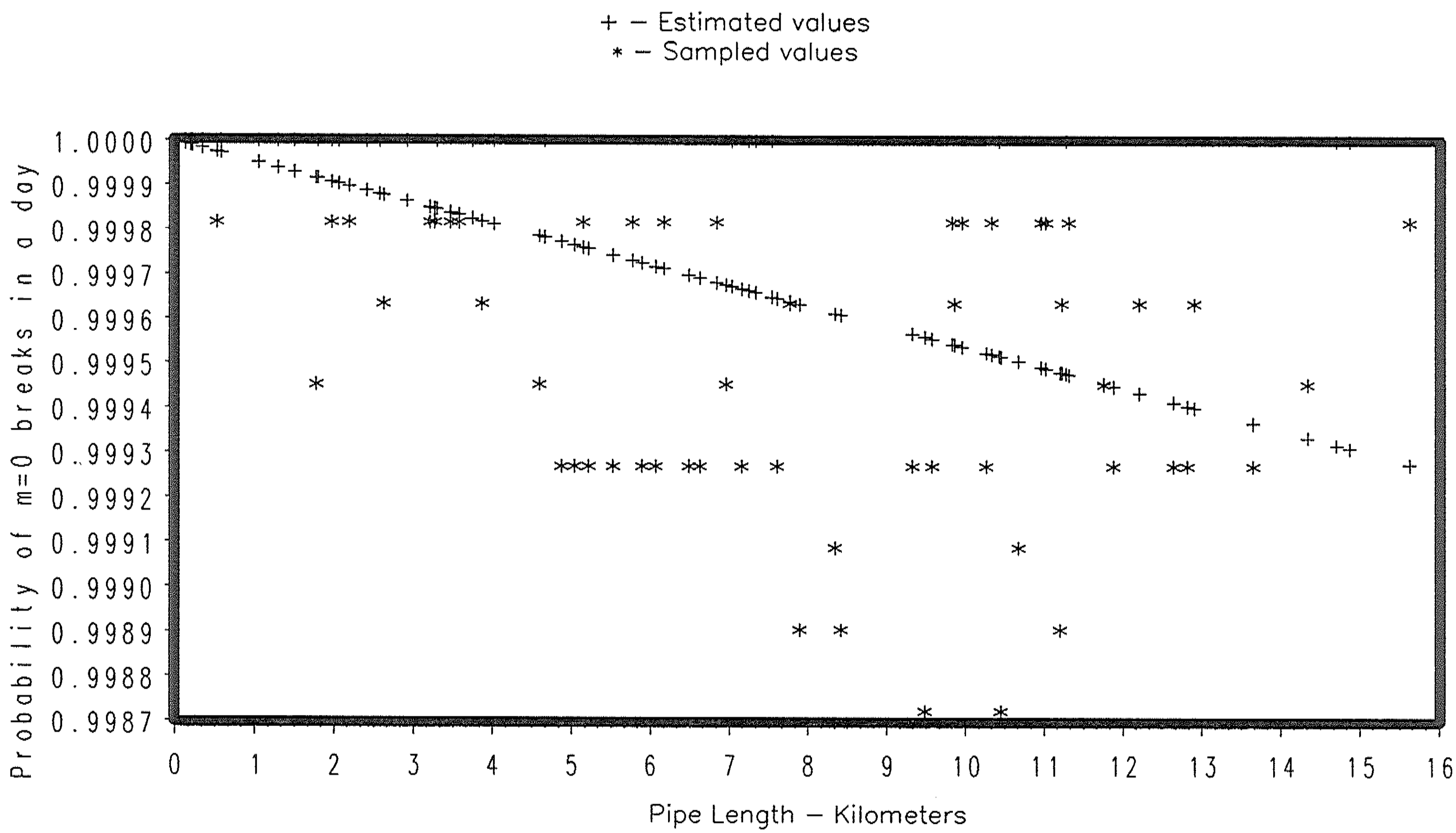


FIGURE B.41 Predicted versus sampled for m=0 breaks: asbestos cement, 250mm 0+ years

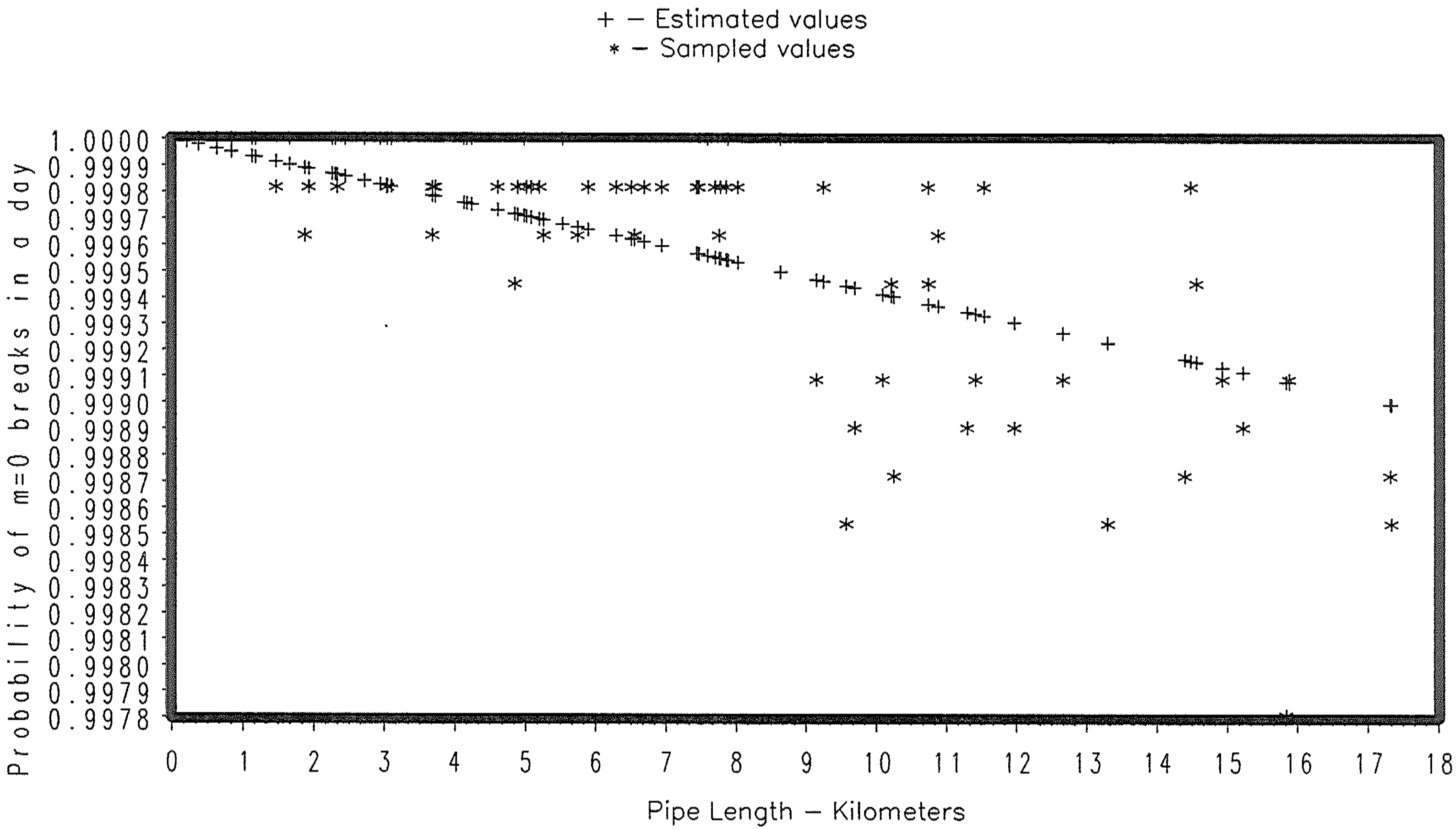


FIGURE B.42 Predicted versus sampled for m=0 breaks: asbestos cement, 300mm 0+ years

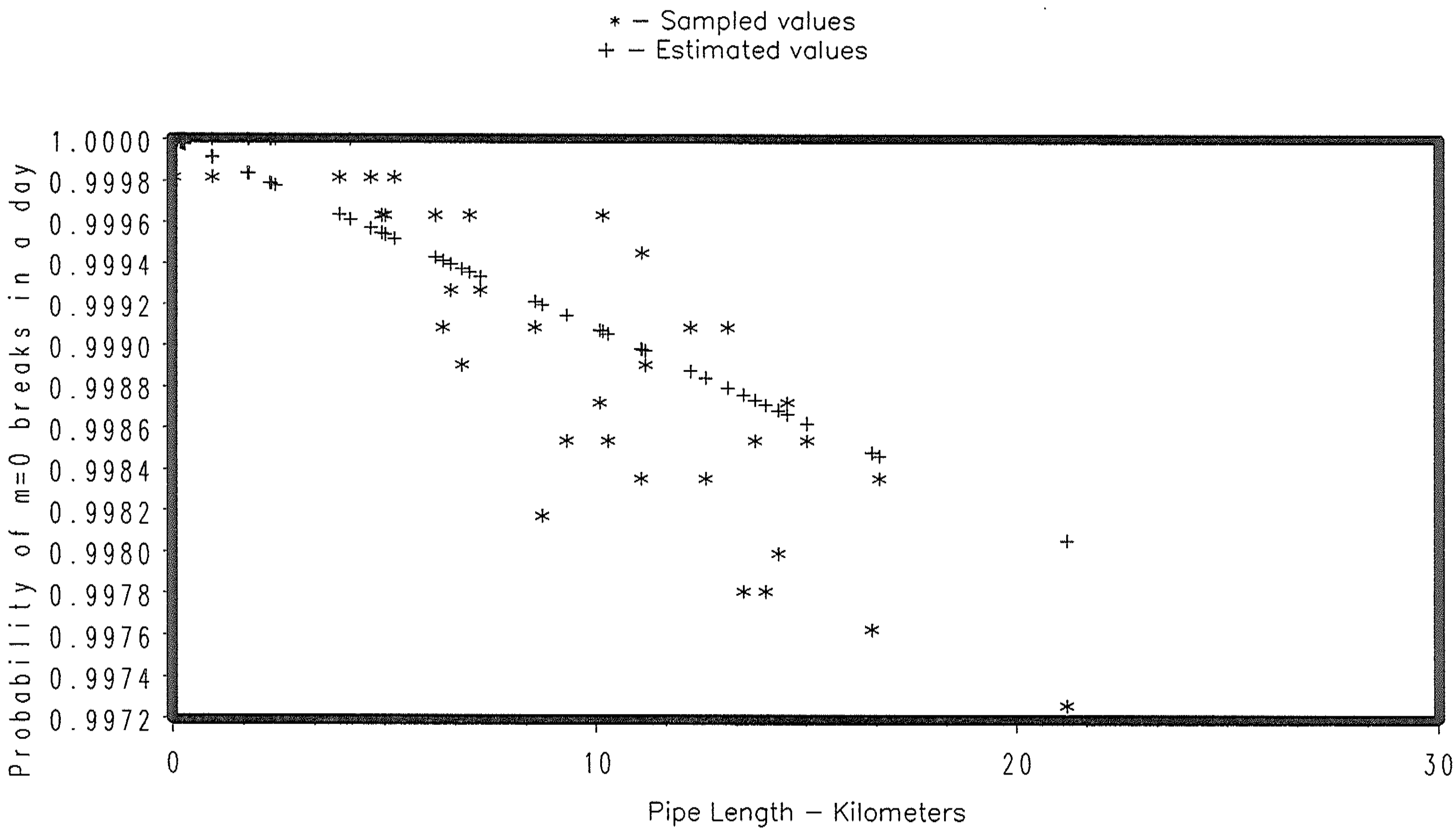


FIGURE B.43 Predicted versus sampled for m=0 breaks: ductile steel, 150 mm 0+ years

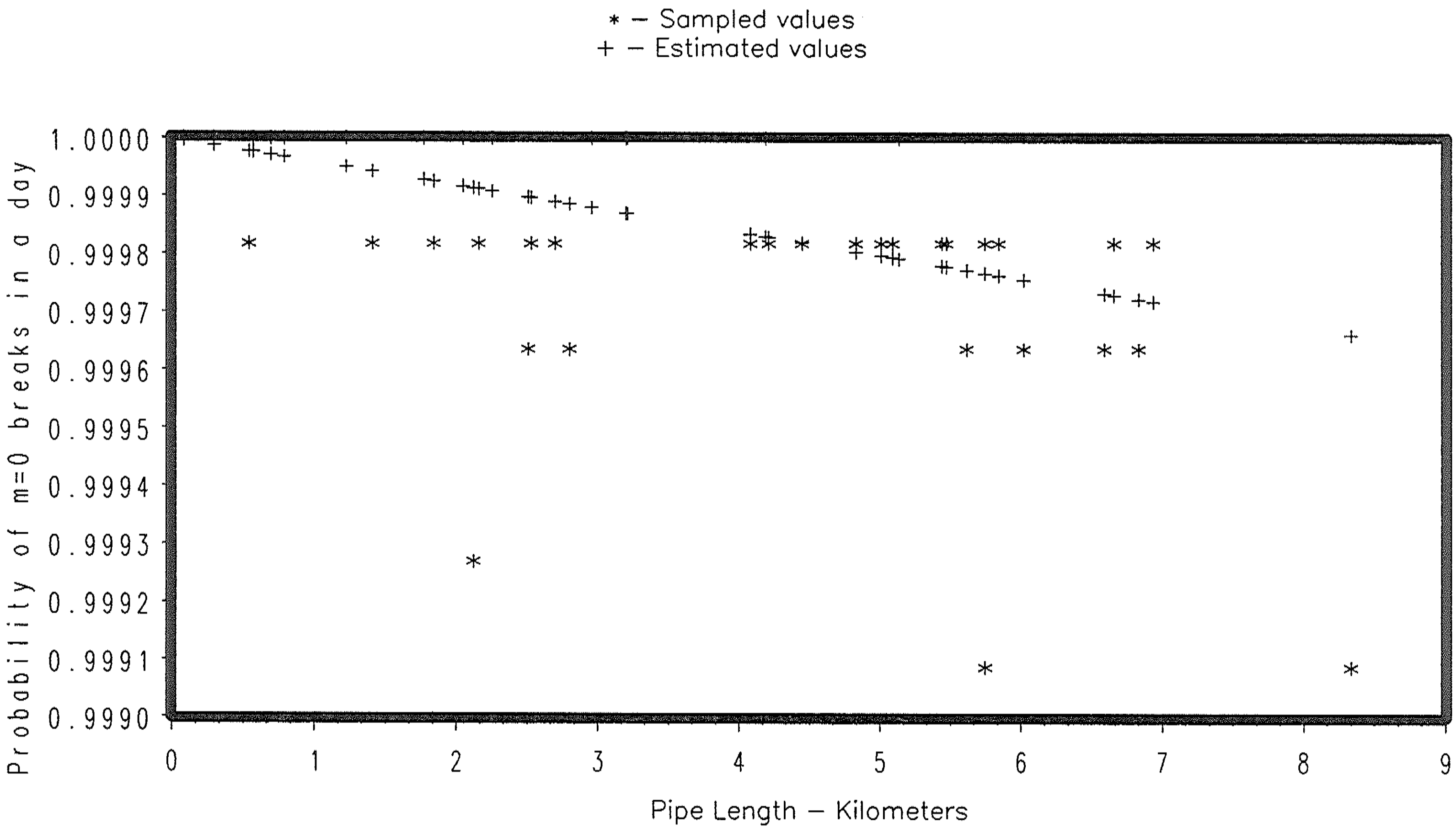


FIGURE B.44 Predicted versus sampled for m=0 breaks: ductile steel, 200 mm 0+ years

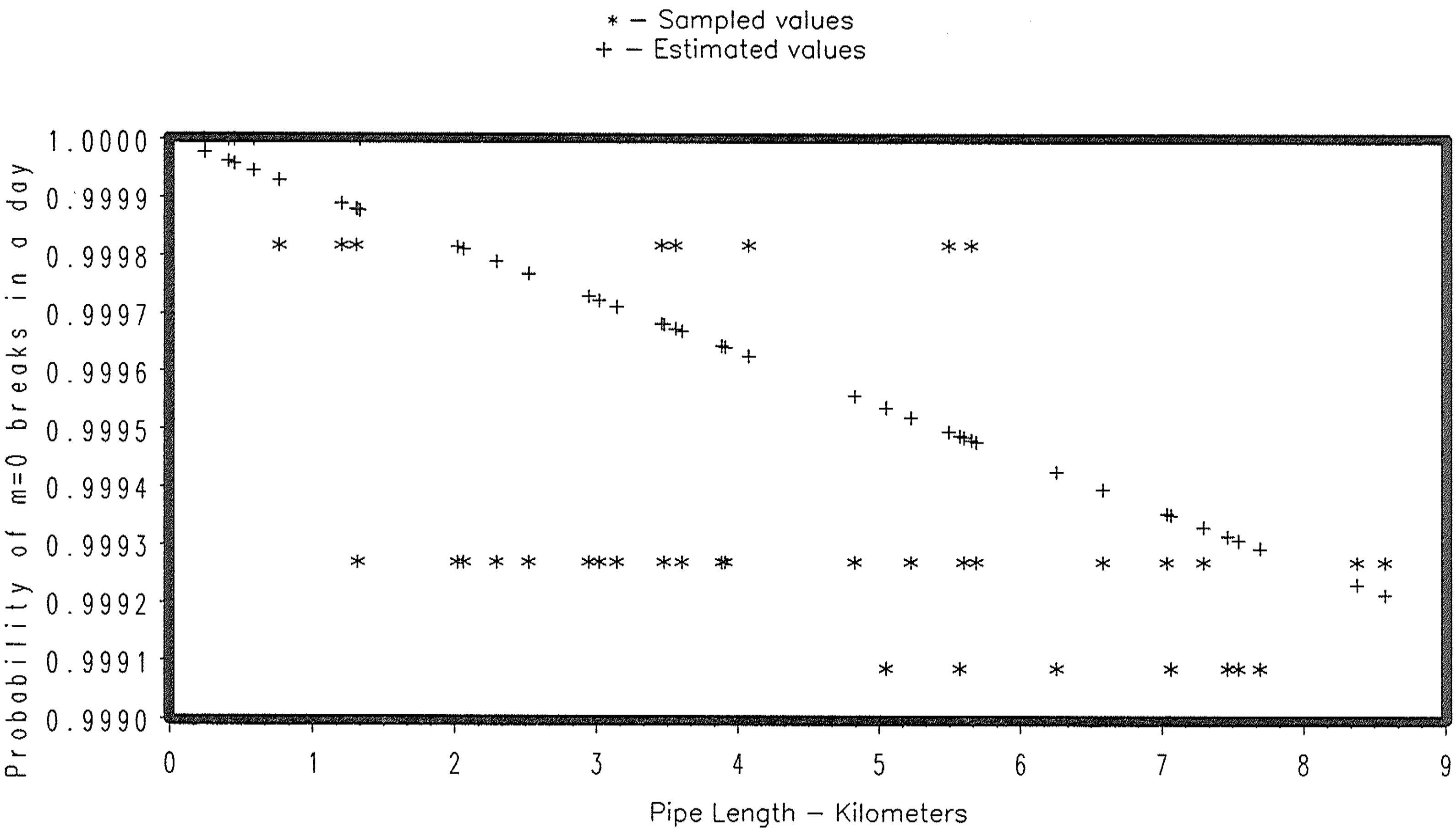


FIGURE B.45 Predicted versus sampled for m=0 breaks: ductile steel, 250 mm 30+ years

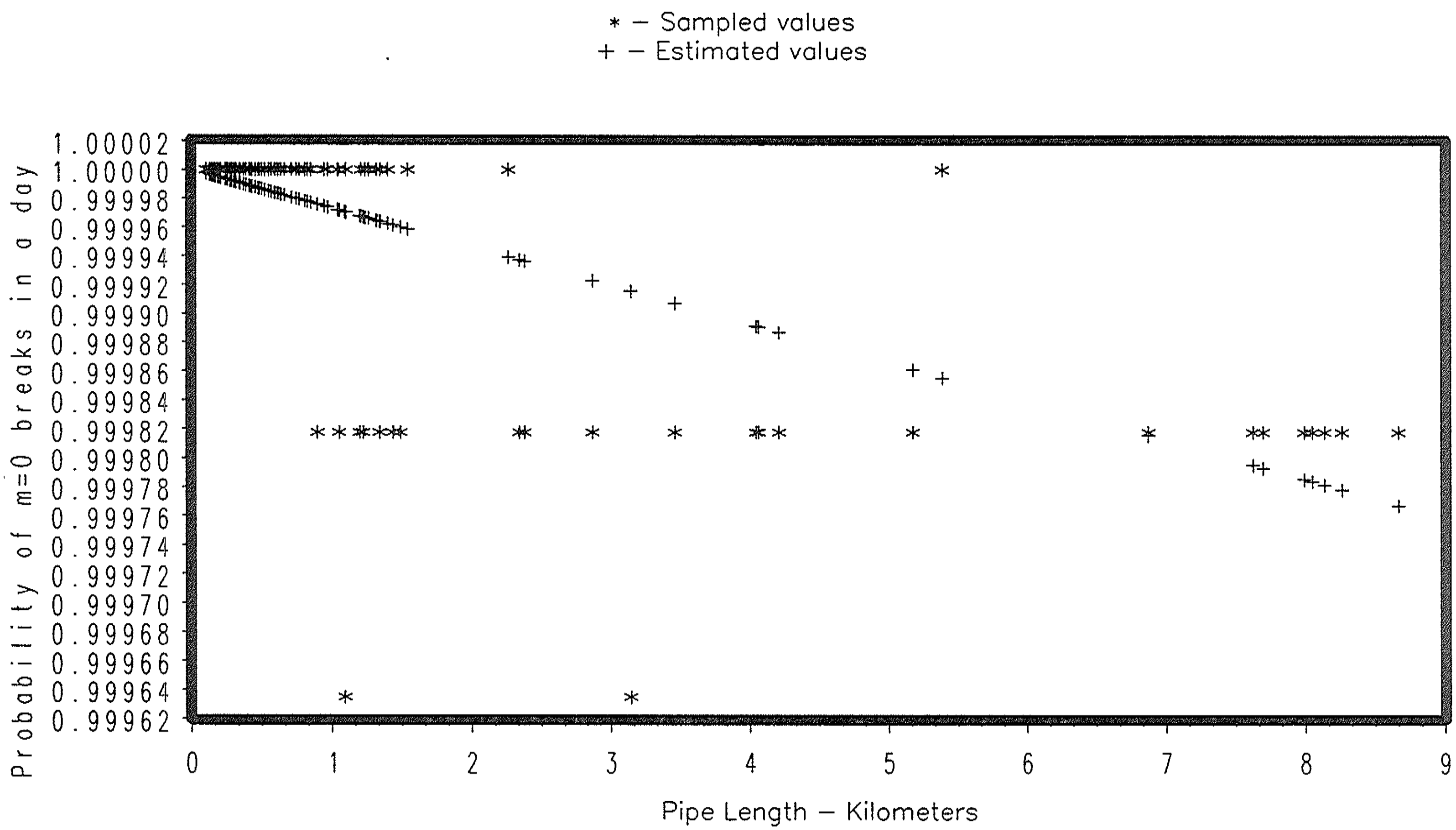


FIGURE B.46 Predicted versus sampled for m=0 breaks: PVC, 150 mm 0+ years

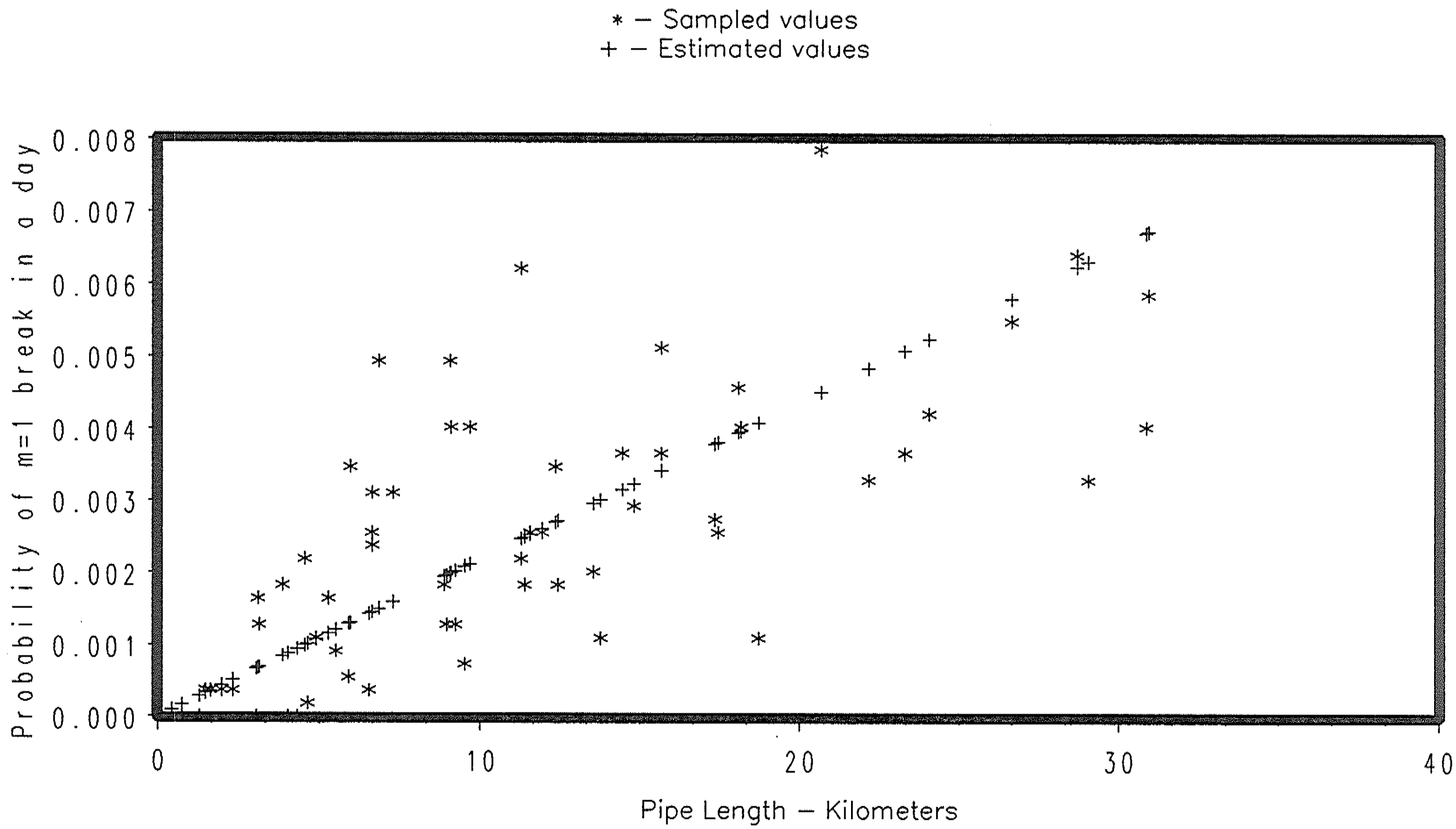


FIGURE B.47 Predicted versus sampled for m=1 break: cast iron pipe, 150mm 0–19 years

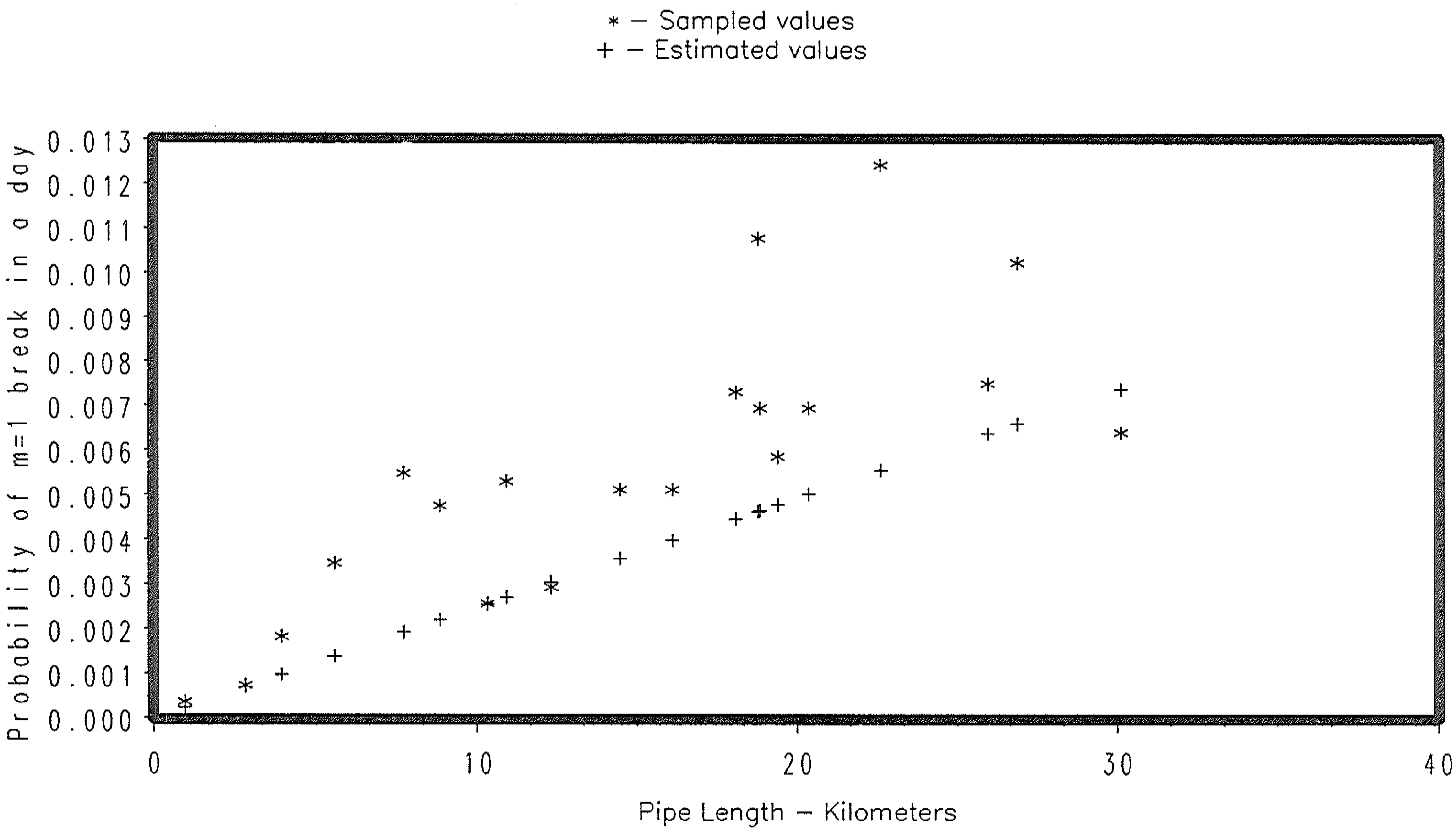


FIGURE B.48 Predicted versus sampled for m=1 break: cast iron pipe, 150mm 20–30 years

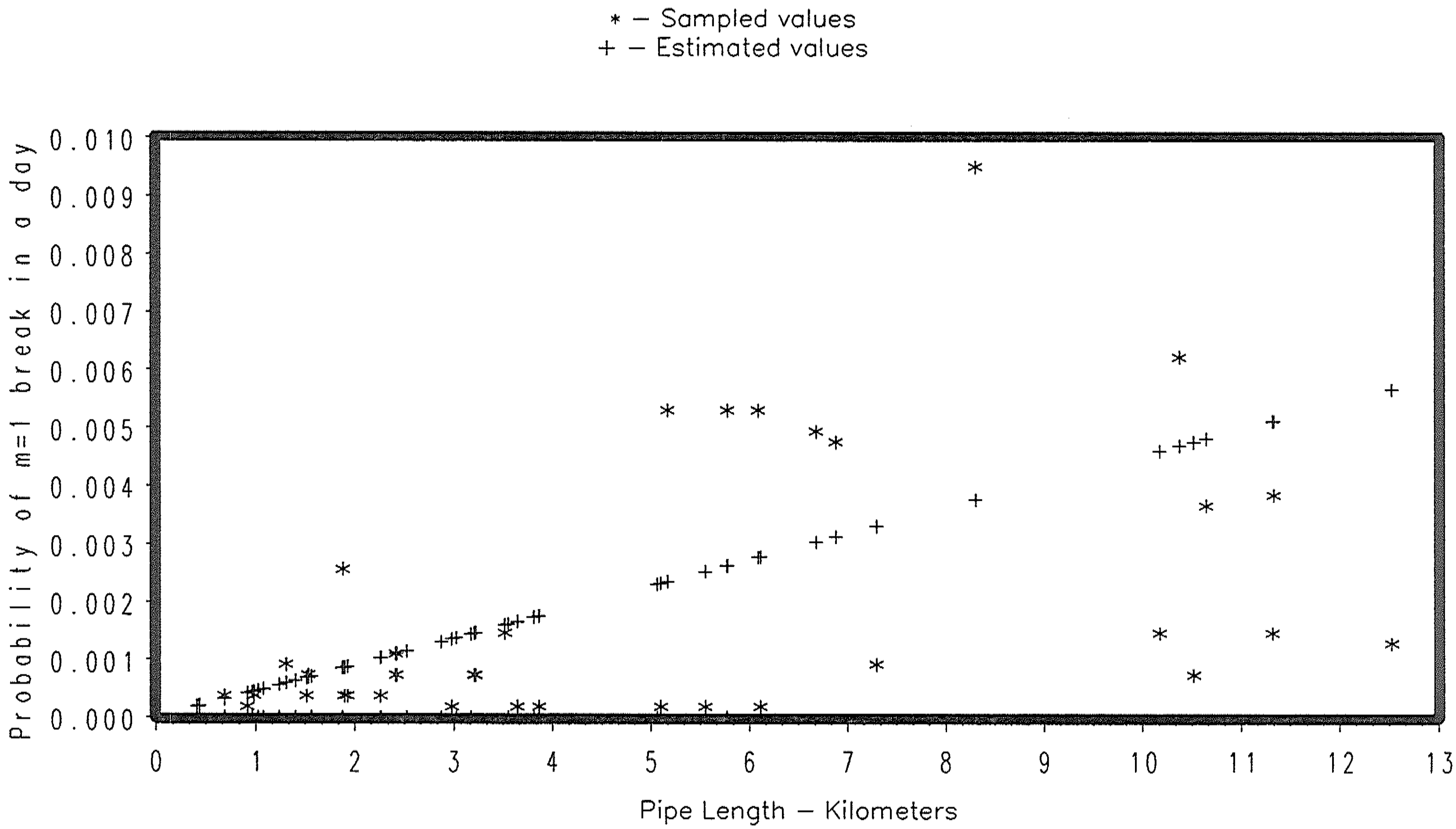


FIGURE B.49 Predicted versus sampled for m=1 break: cast iron pipe, 150mm 31+ years

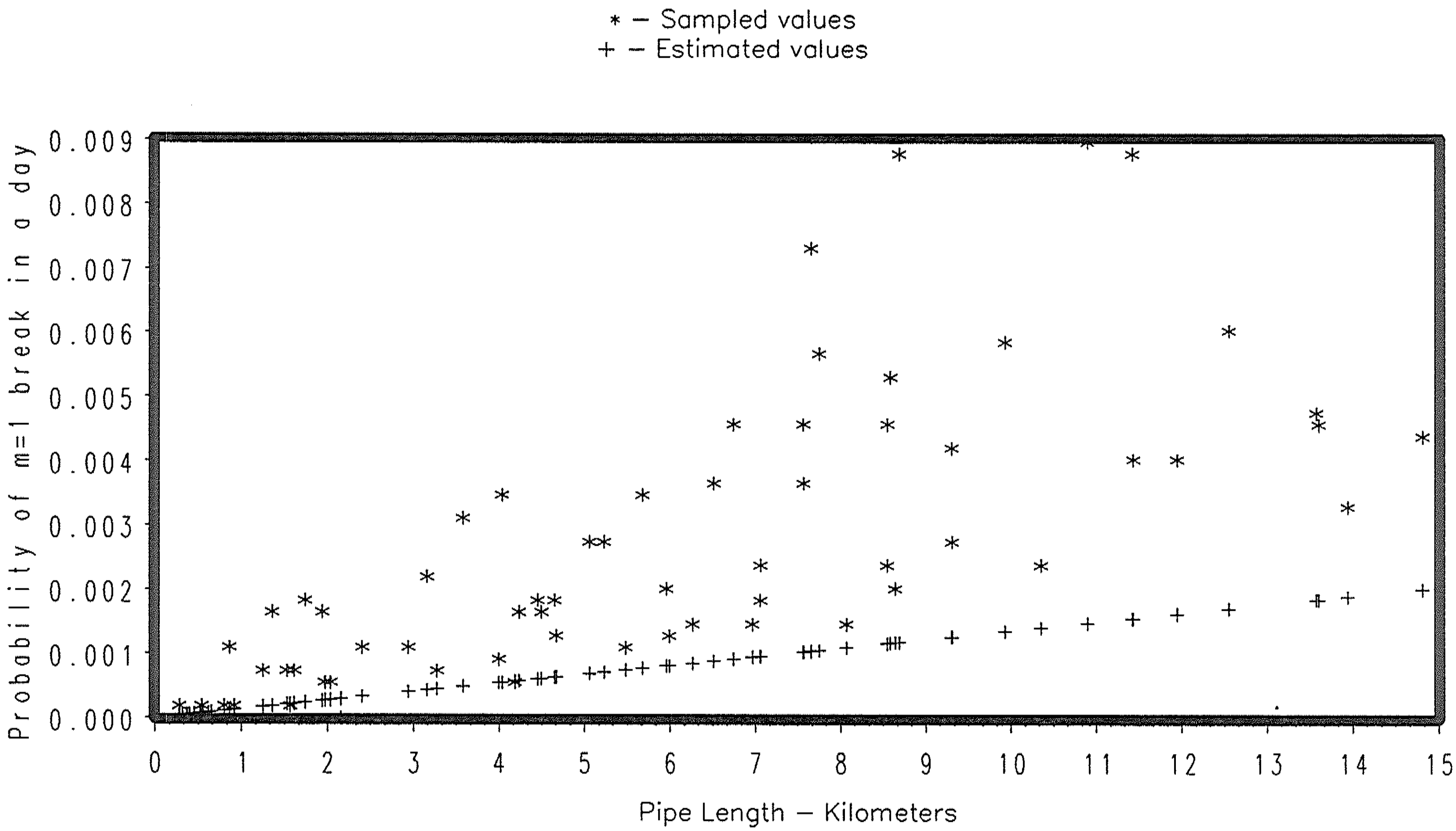


FIGURE B.50 Predicted versus sampled for m=1 break: cast iron pipe, 200mm 0–30 years

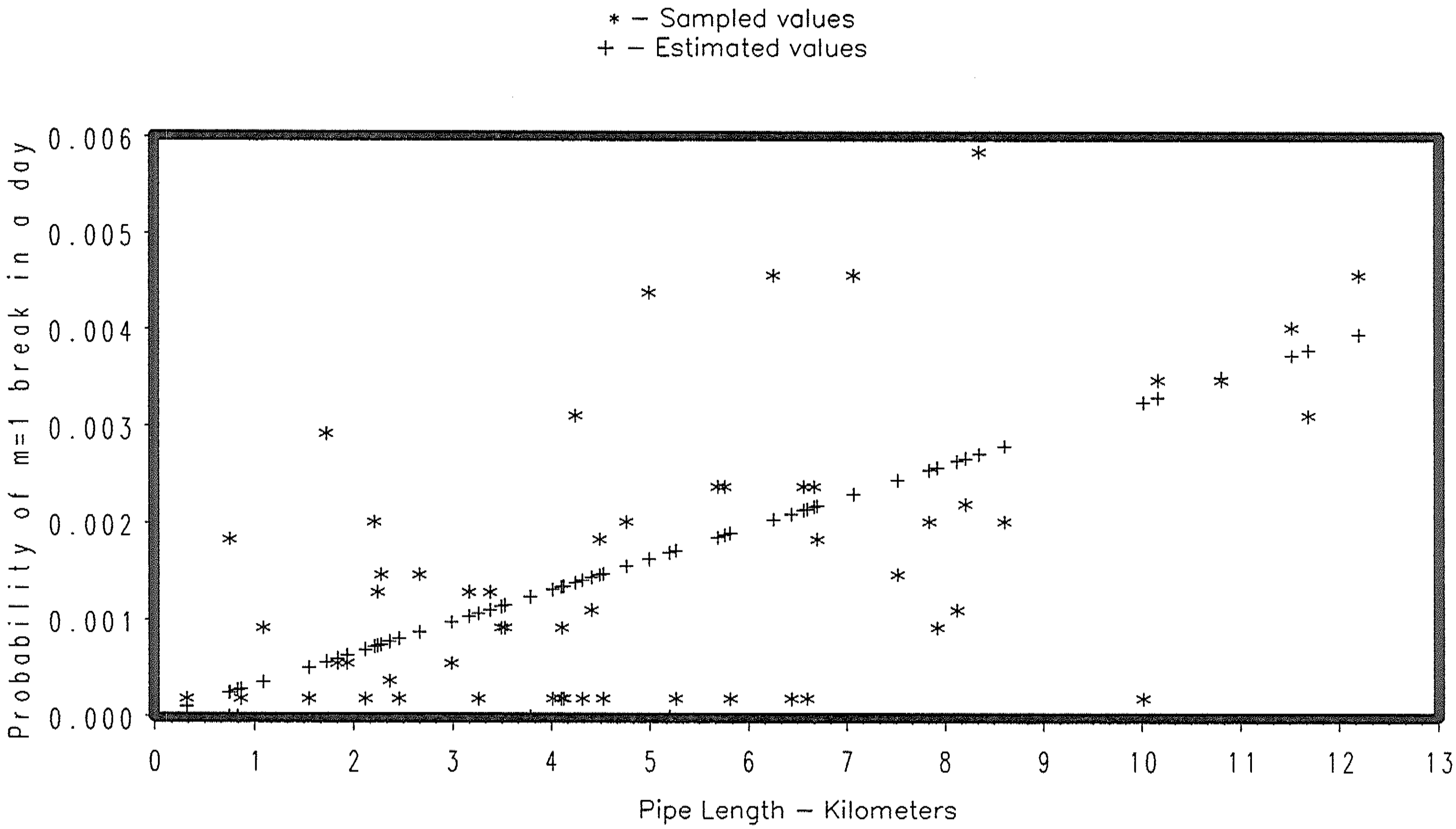


FIGURE B.51 Predicted versus sampled for m=1 break: cast iron pipe, 200mm 31+ years

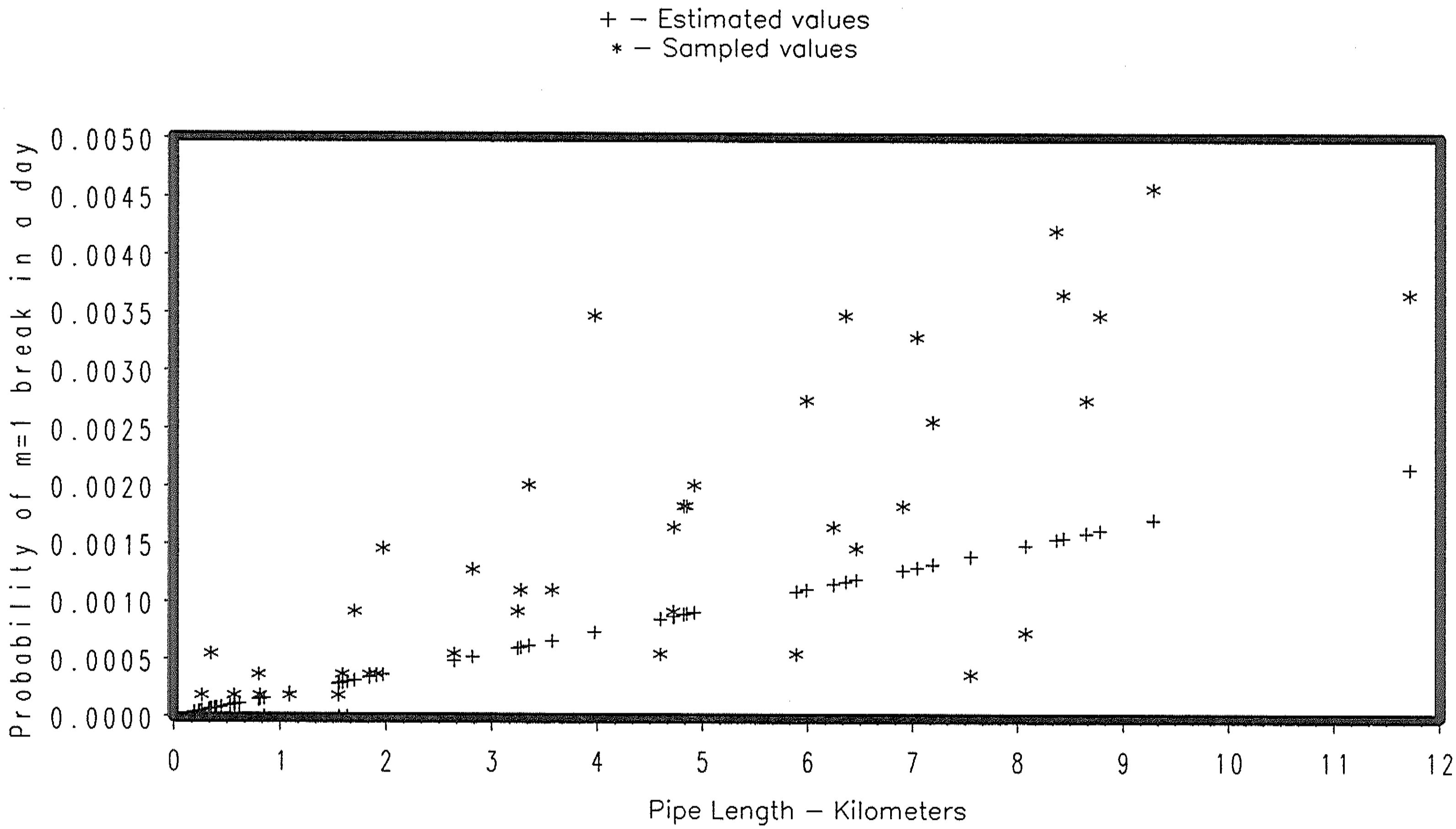


FIGURE B.52 Predicted versus sampled for m=1 break: cast iron pipe, 250mm 0–30 years

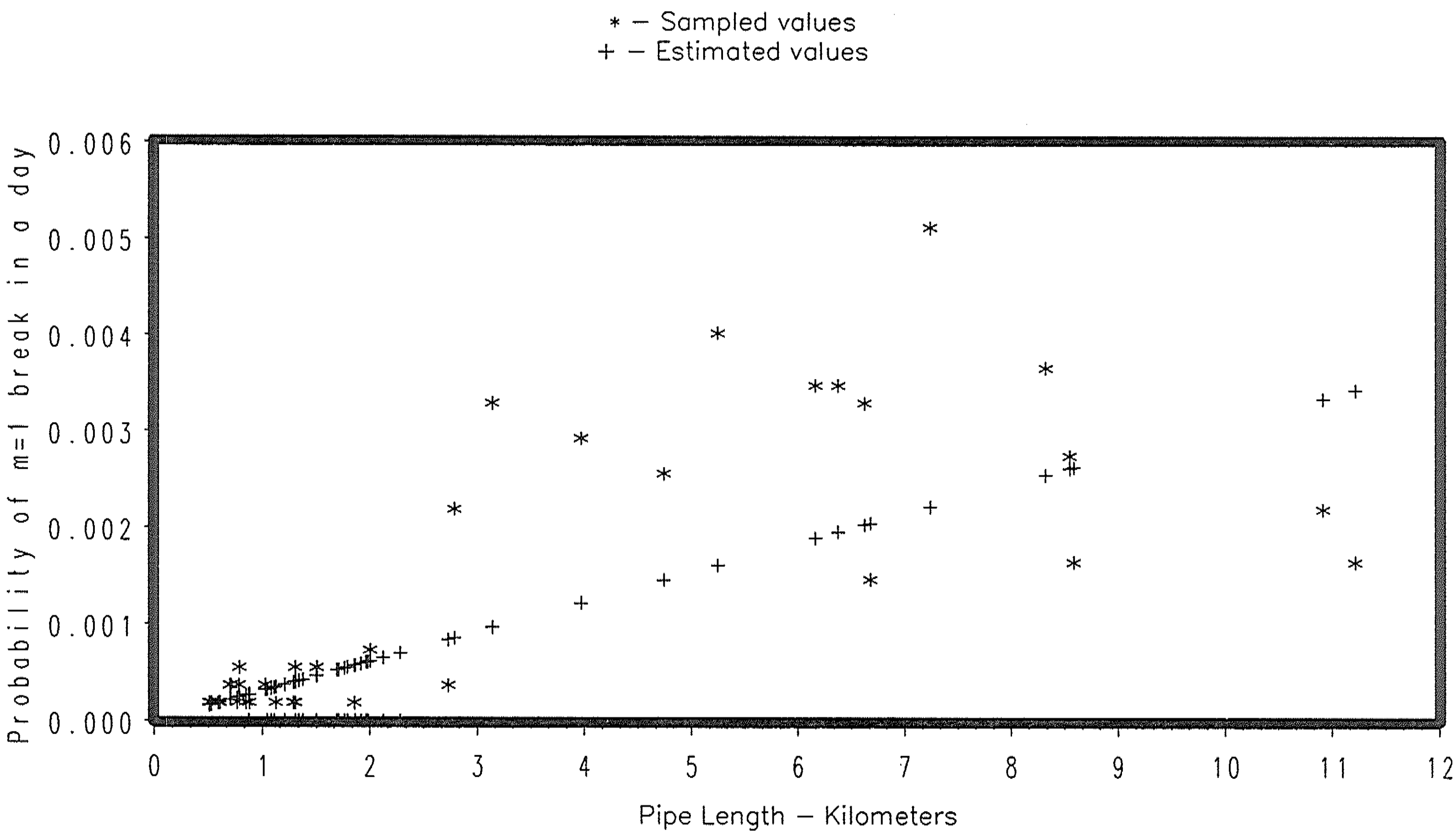


FIGURE B.53 Predicted versus sampled for m=1 break: cast iron pipe, 250mm 31+ years

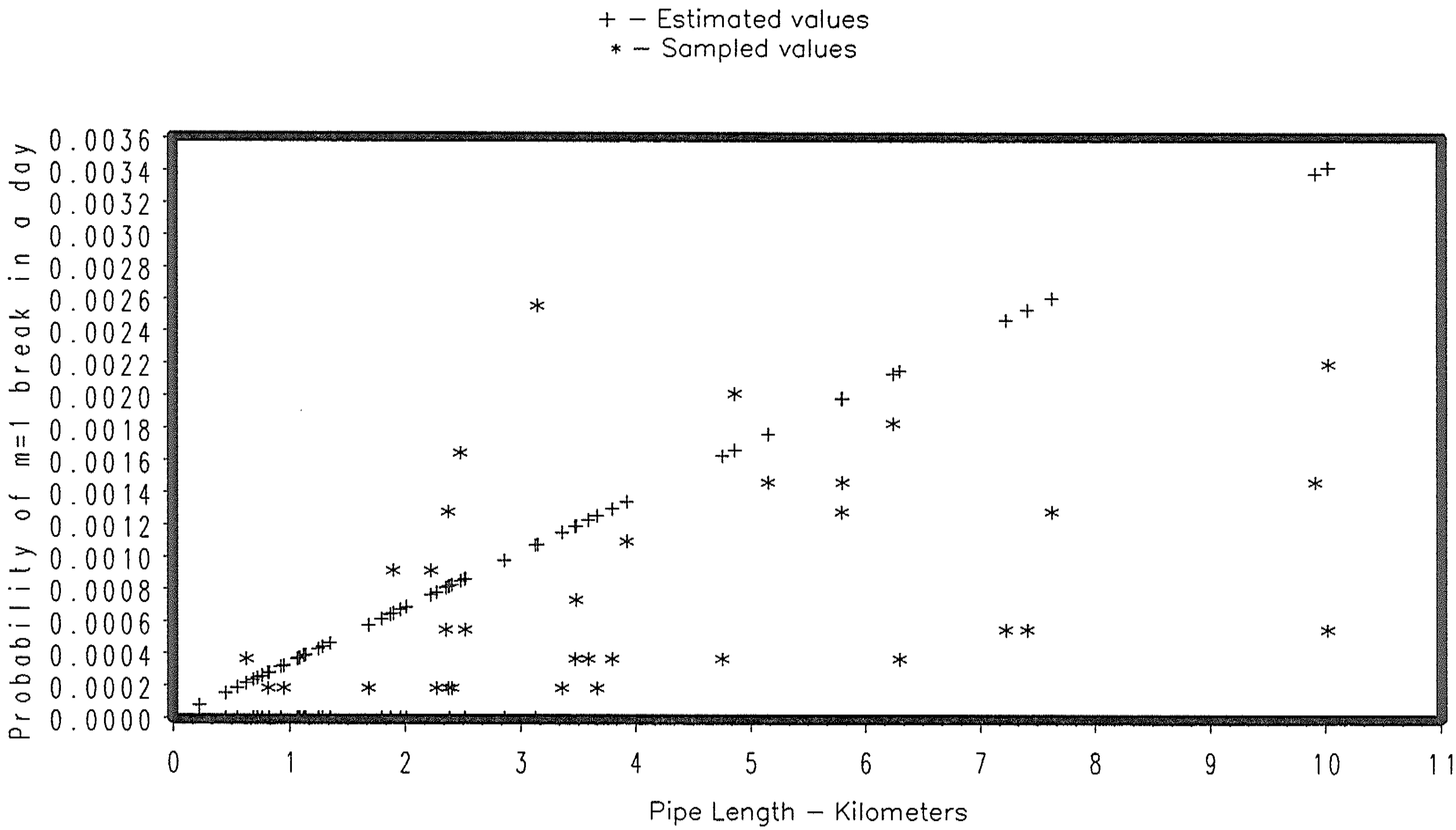


FIGURE B.54 Predicted versus sampled for m=1 break: cast iron pipe, 300mm 0–30 years

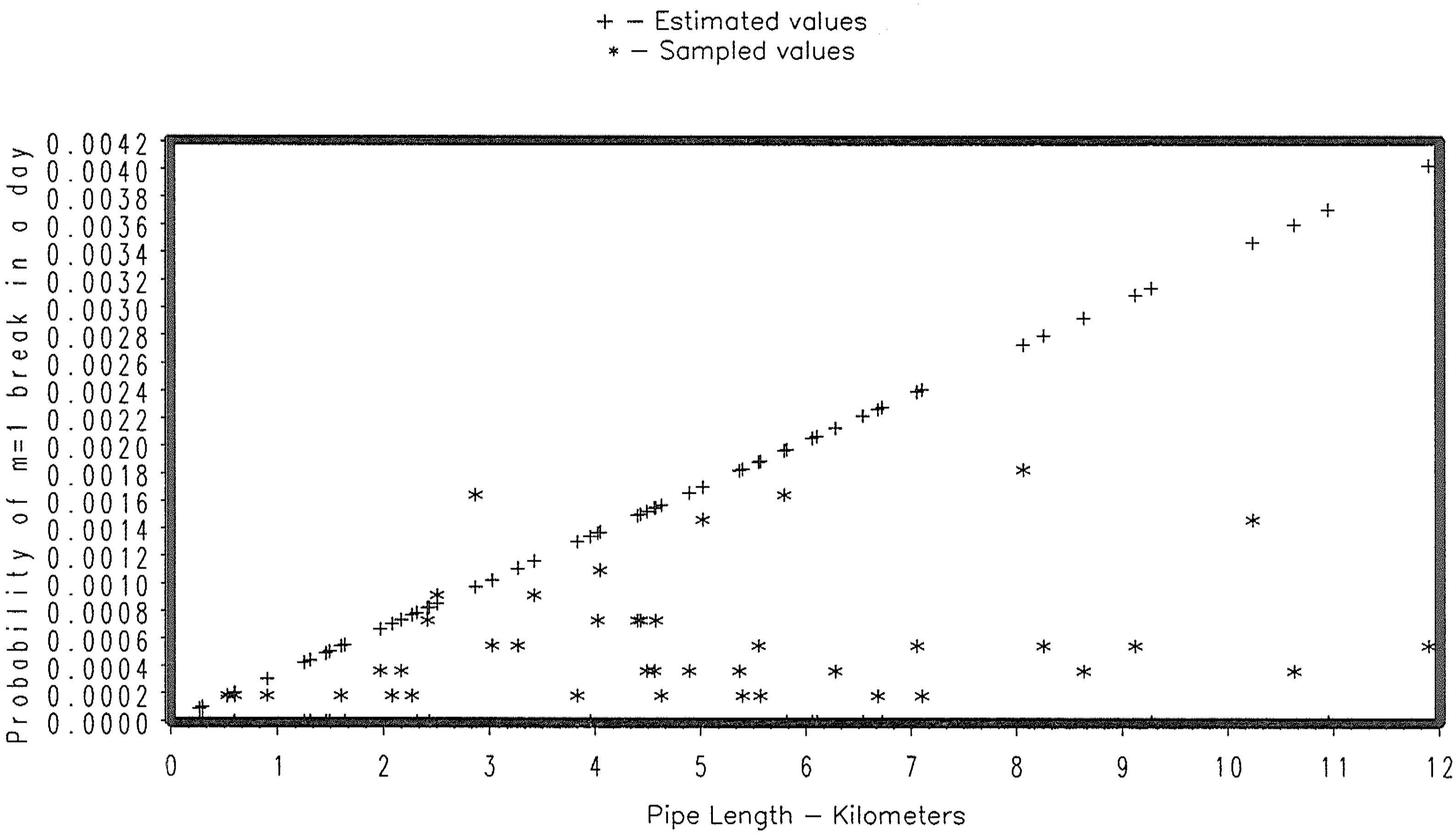


FIGURE B.55 Predicted versus sampled for m=1 break: cast iron pipe, 300mm 31+ years

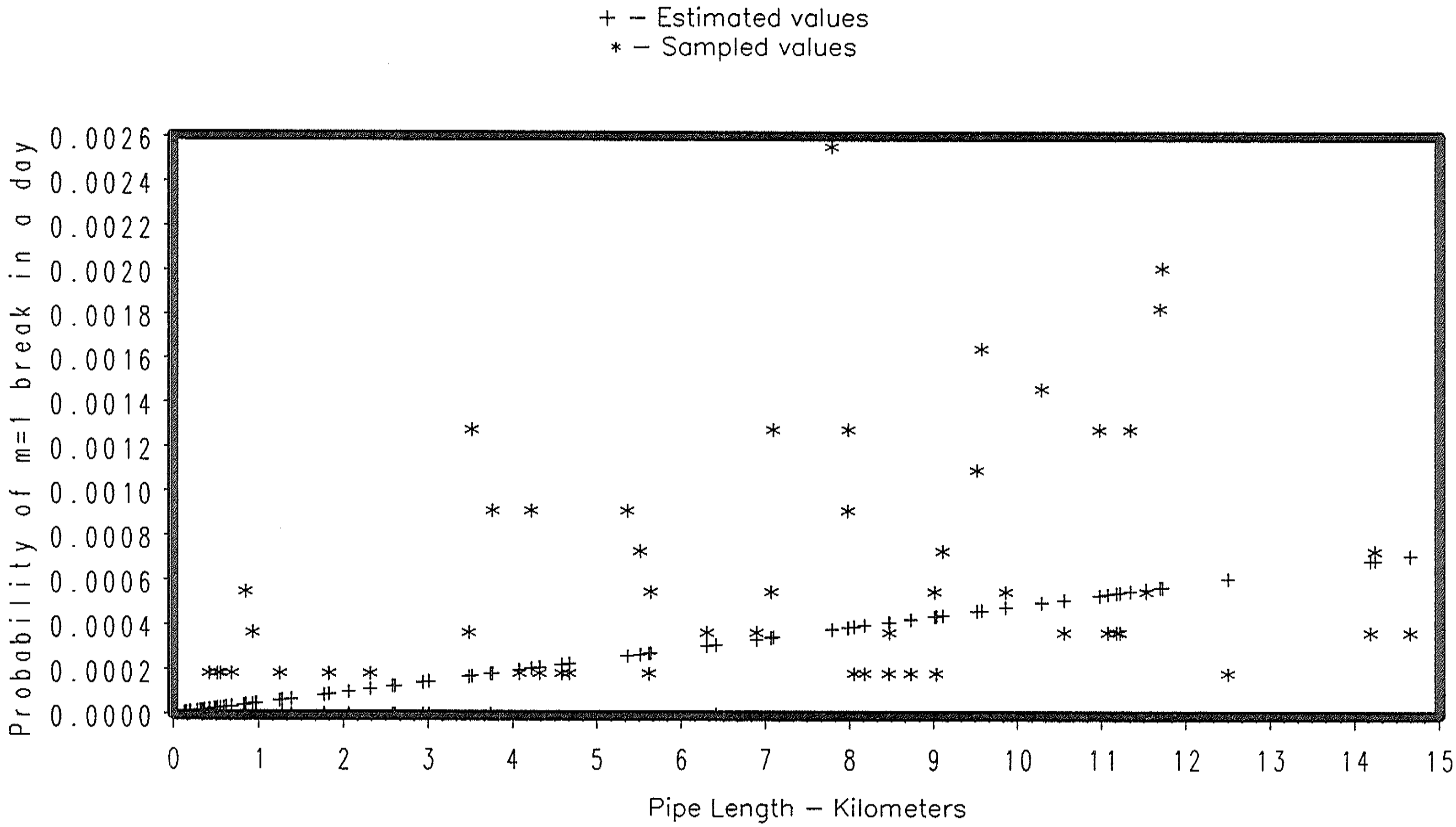


FIGURE B.56 Predicted versus sampled for m=1 break: asbestos cement, 150mm 0–30 years

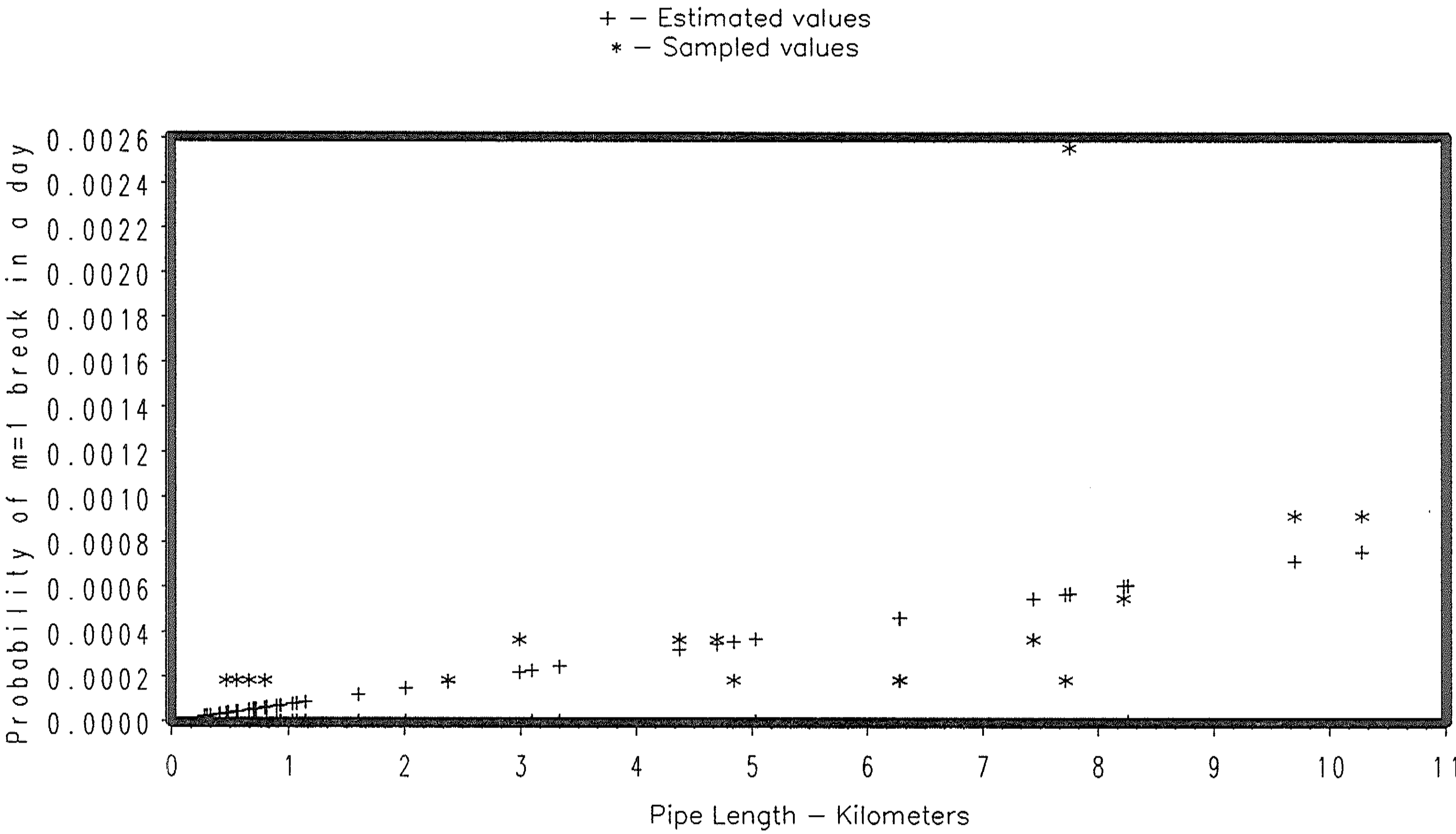


FIGURE B.57 Predicted versus sampled for m=1 break: asbestos cement,150mm 31+ years

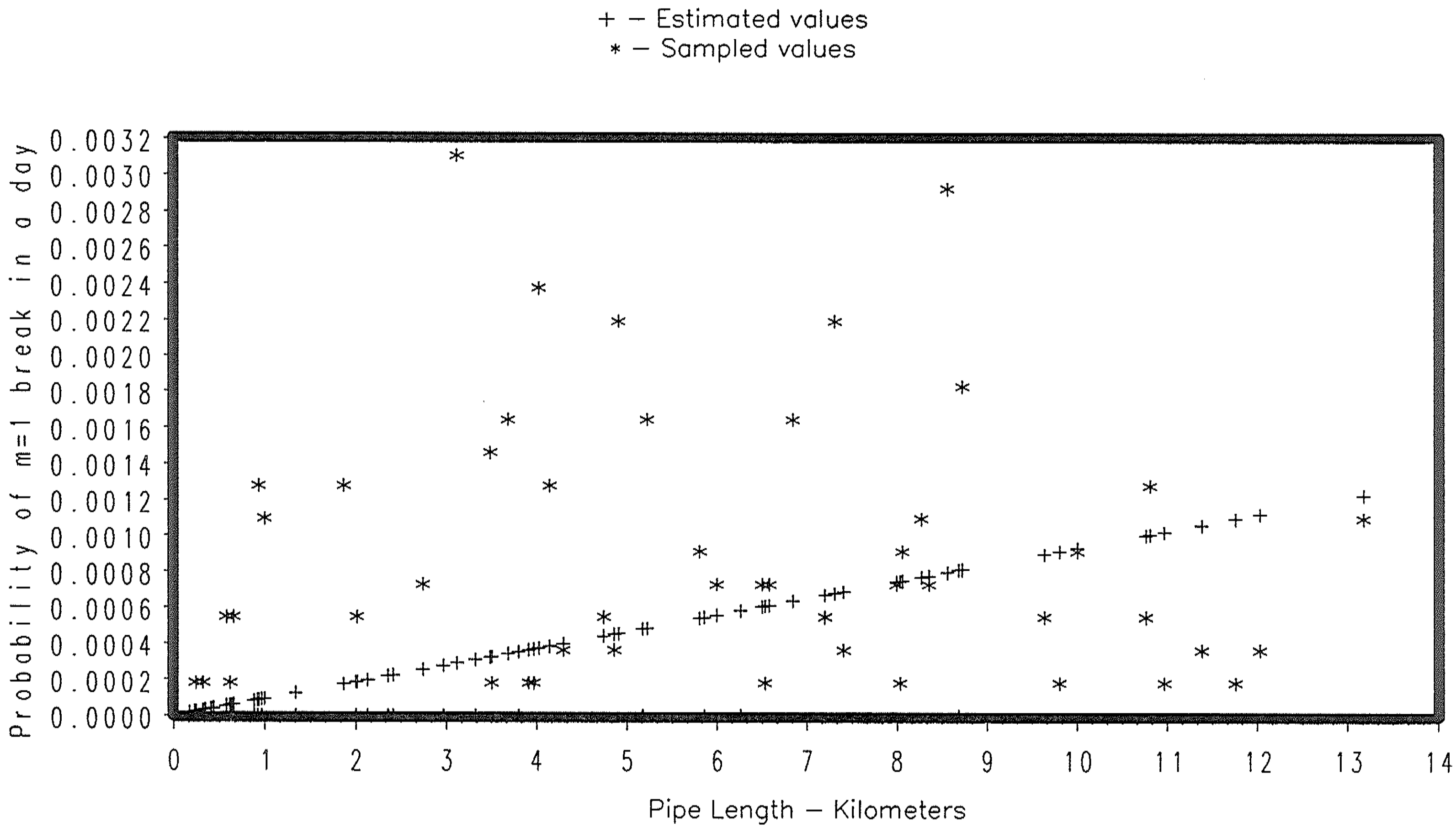


FIGURE B.58 Predicted versus sampled for m=1 break: asbestos cement, 200mm 0–30 years

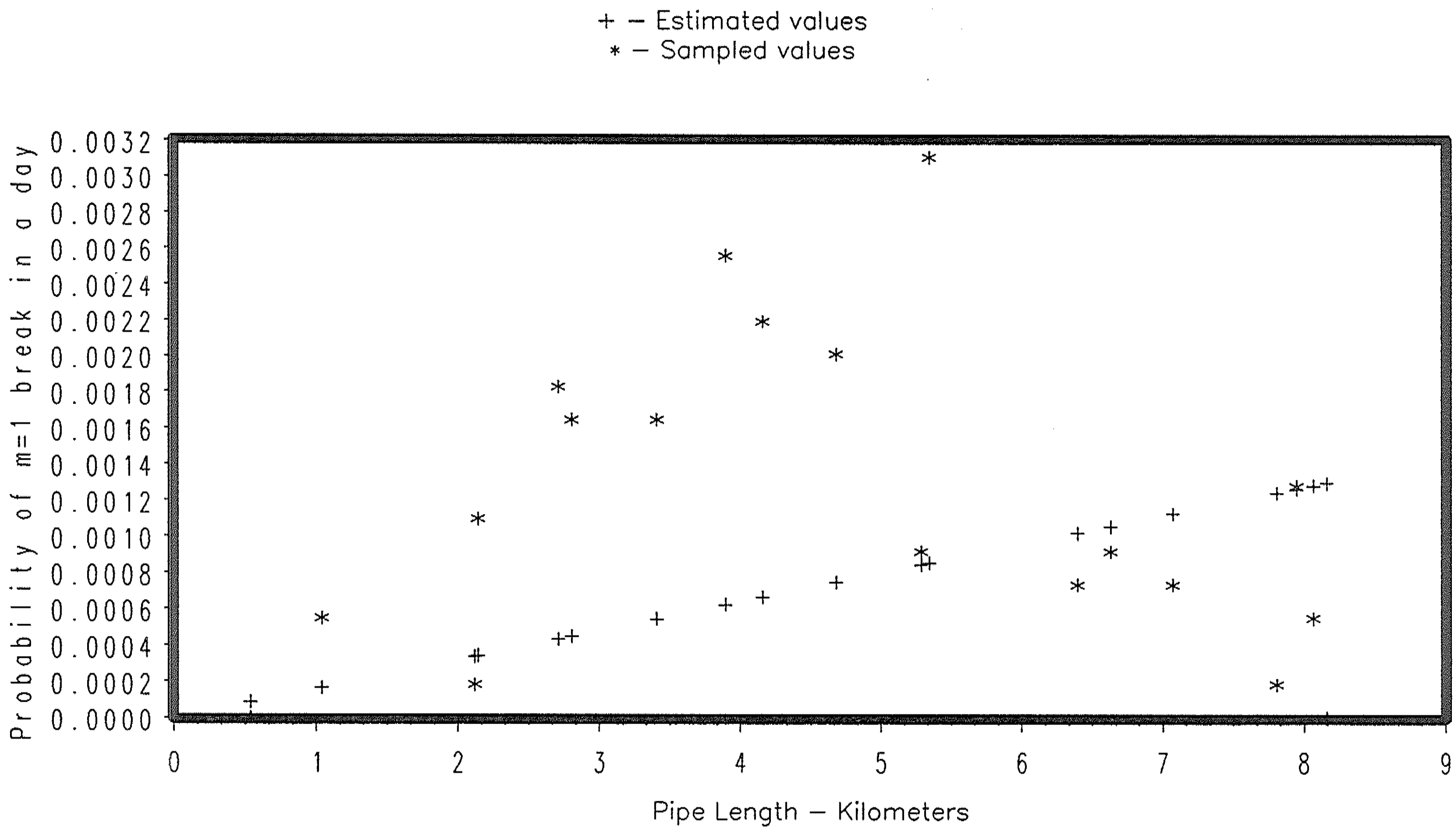


FIGURE B.59 Predicted versus sampled for m=1 break: asbestos cement, 200mm 31+ years

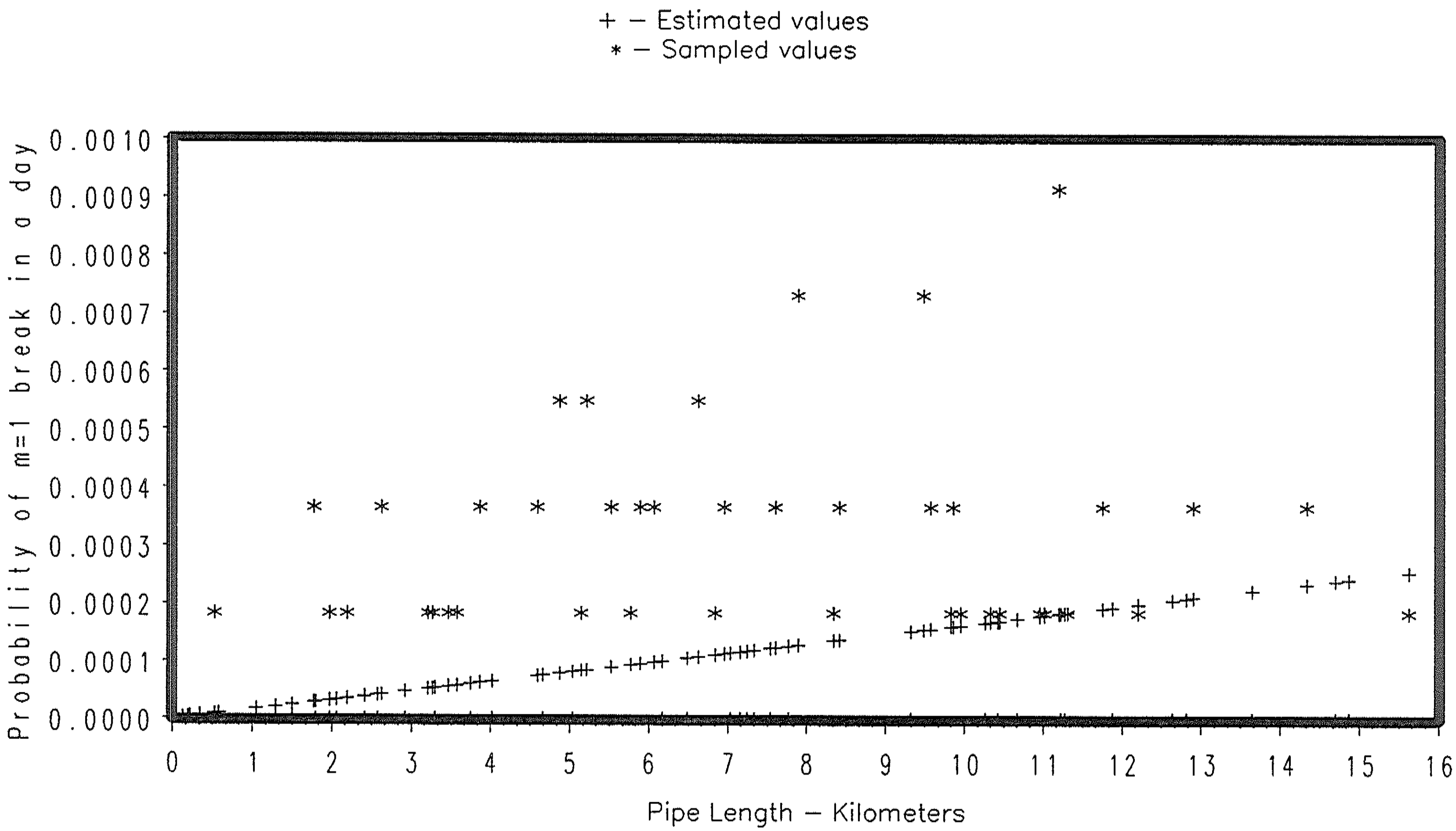


FIGURE B.60 Predicted versus sampled for m=1 break: asbestos cement, 250mm 0+ years

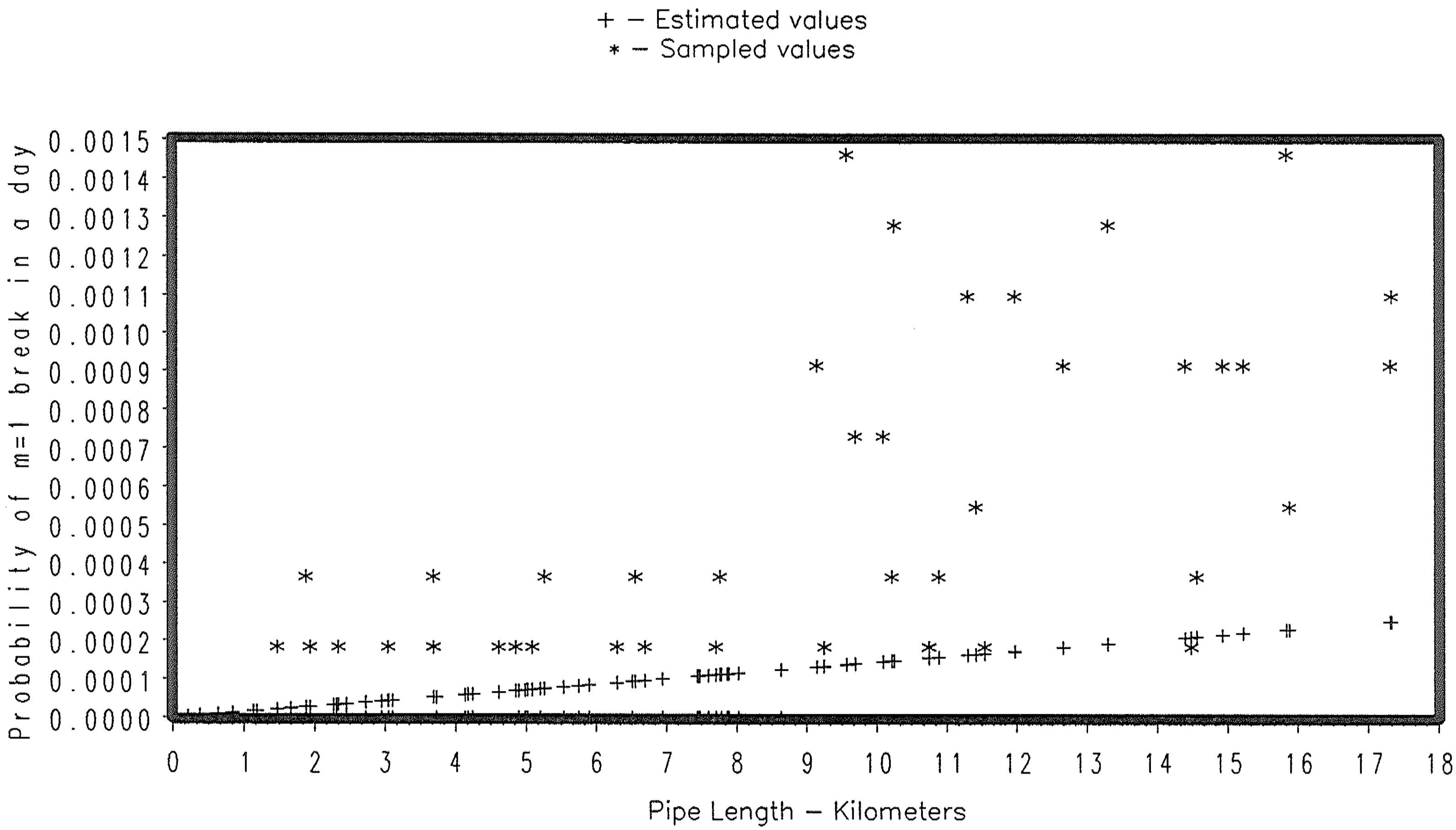


FIGURE B.61 Predicted versus sampled for m=1 break: asbestos cement, 300mm 0+ years

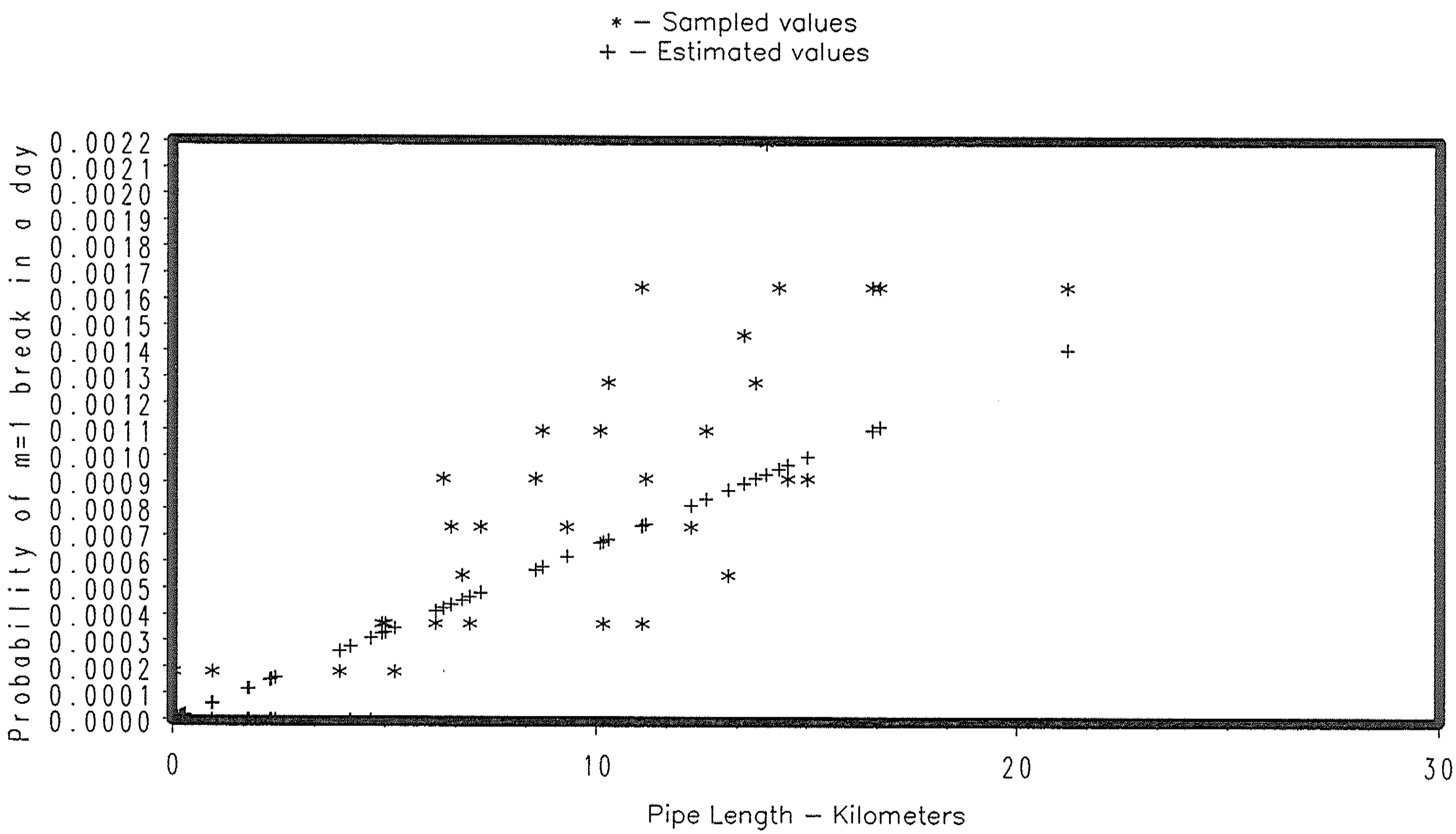


FIGURE B.62 Predicted versus sampled for m=1 break: ductile steel pipe, 150mm 0+ years

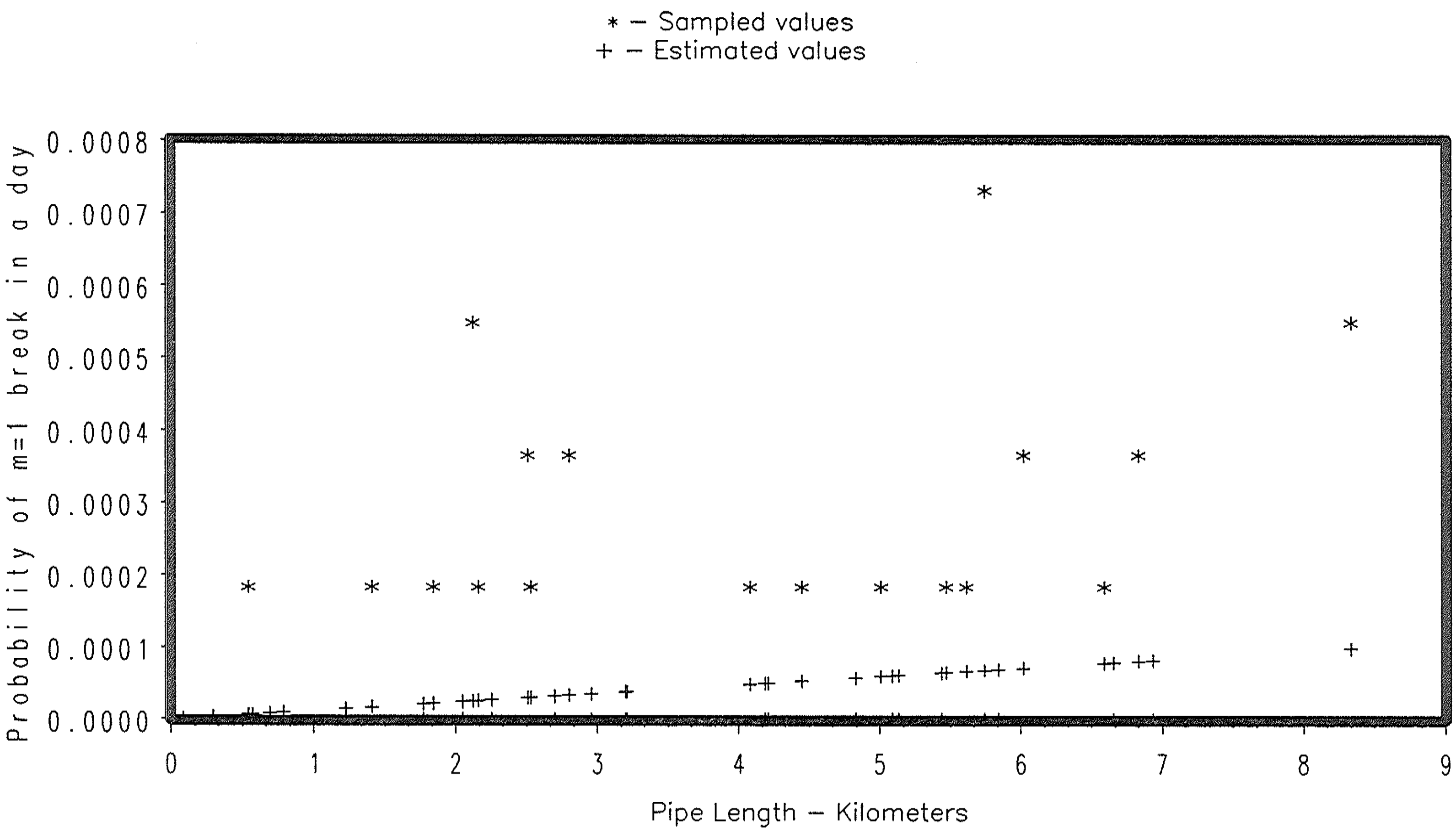


FIGURE B.63 Predicted versus sampled for m=1 break: ductile steel pipe, 200mm 0+ years

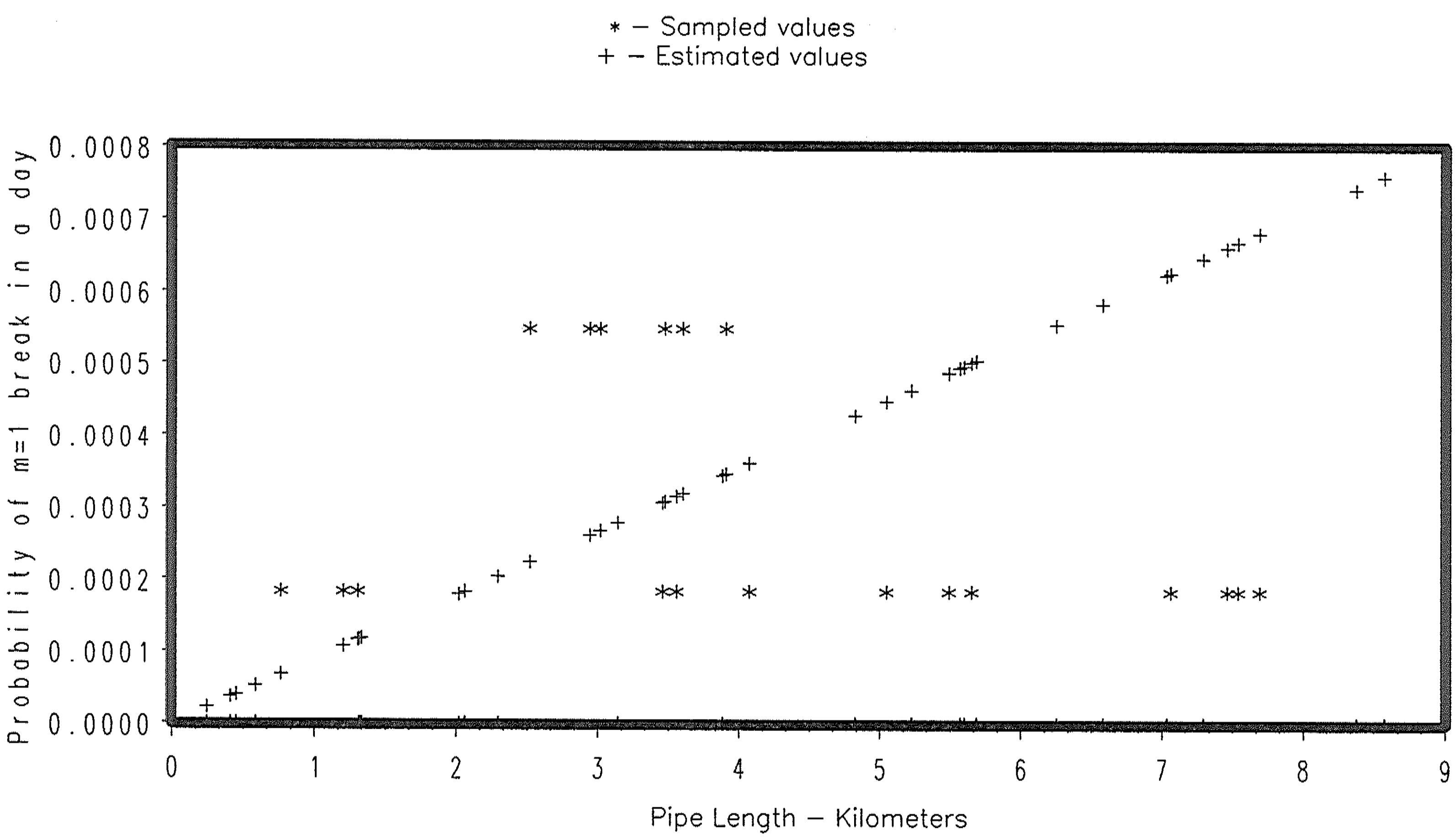


FIGURE B.64 Predicted versus sampled for m=1 break: ductile steel, 250mm 0+ years

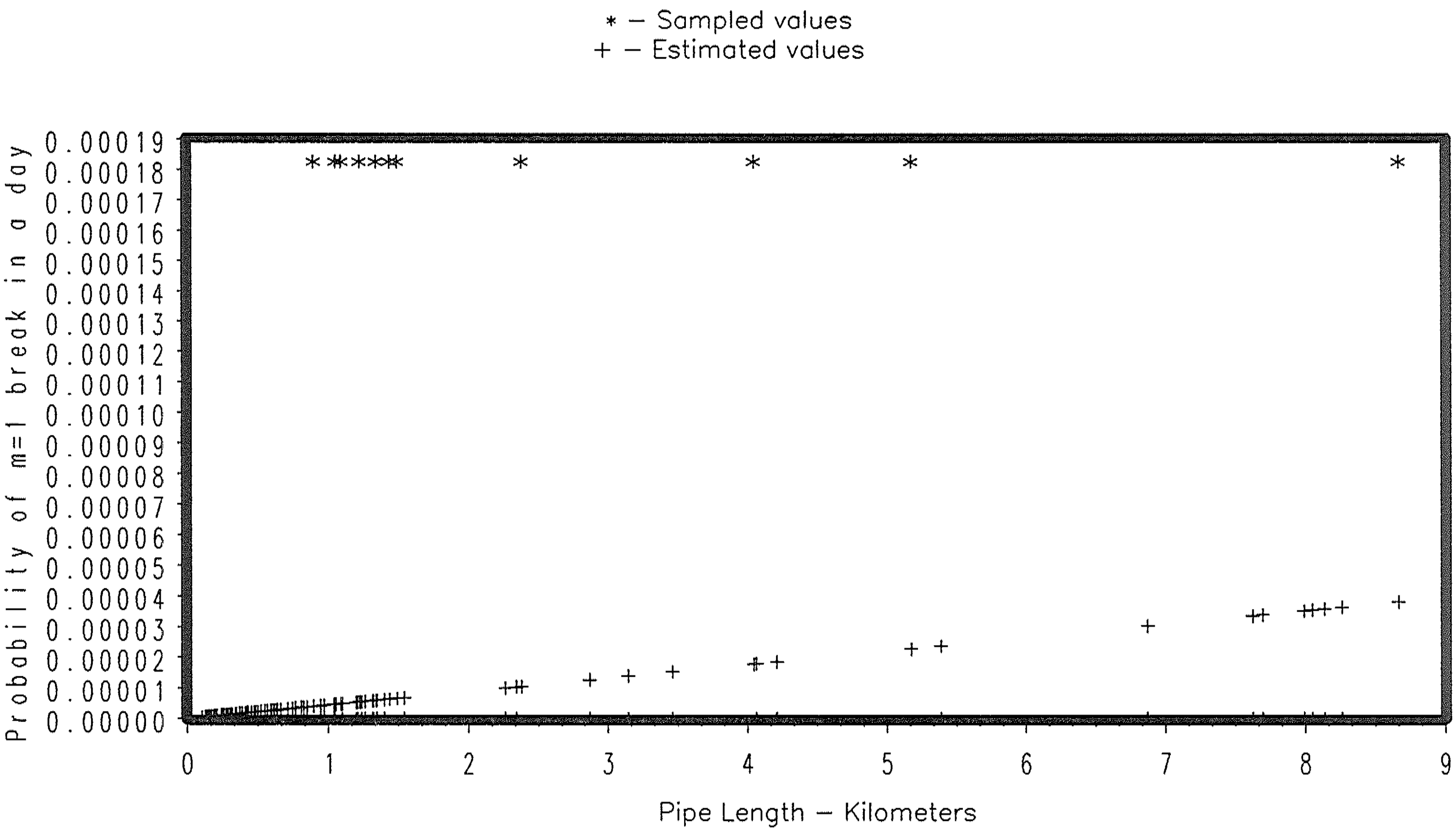


FIGURE B.65 Predicted versus sampled for m=1 break: PVC pipe, 150mm 0+ years

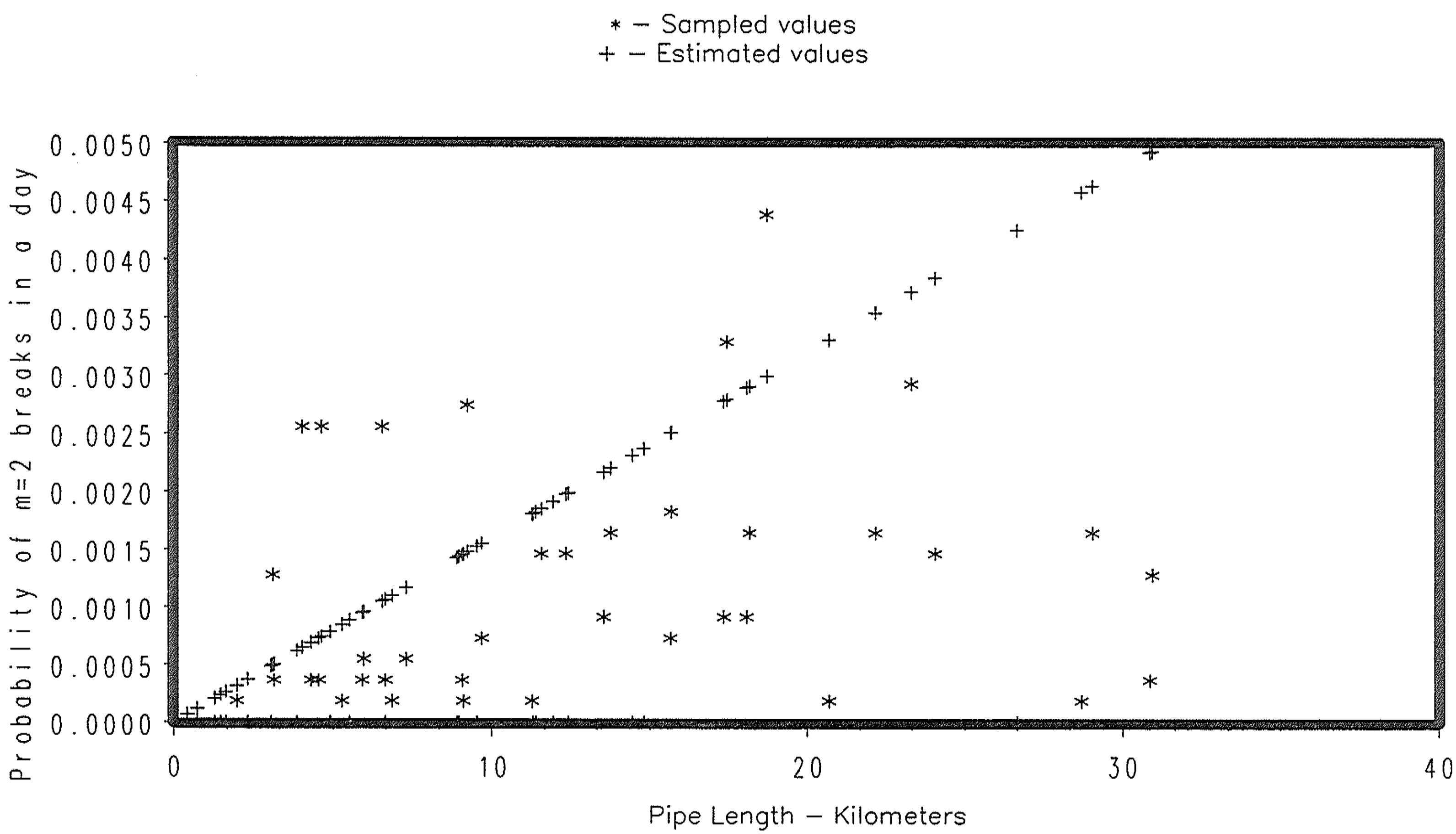


FIGURE B.66 Predicted versus sampled for m=2 breaks: cast iron pipe, 150mm 0–19 years

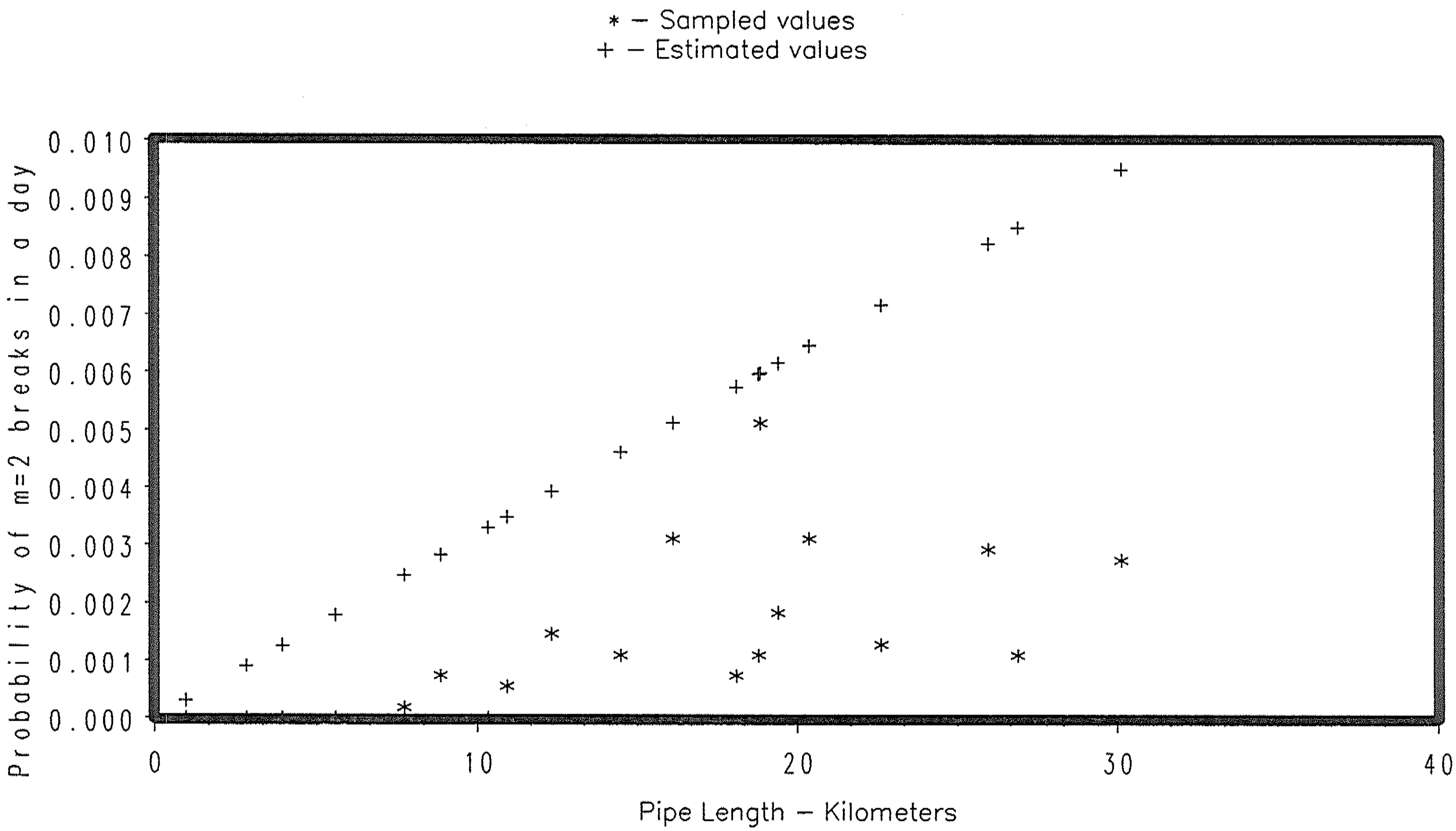


FIGURE B.67 Predicted versus sampled for m=2 breaks: cast iron pipe, 150mm 20–30 years

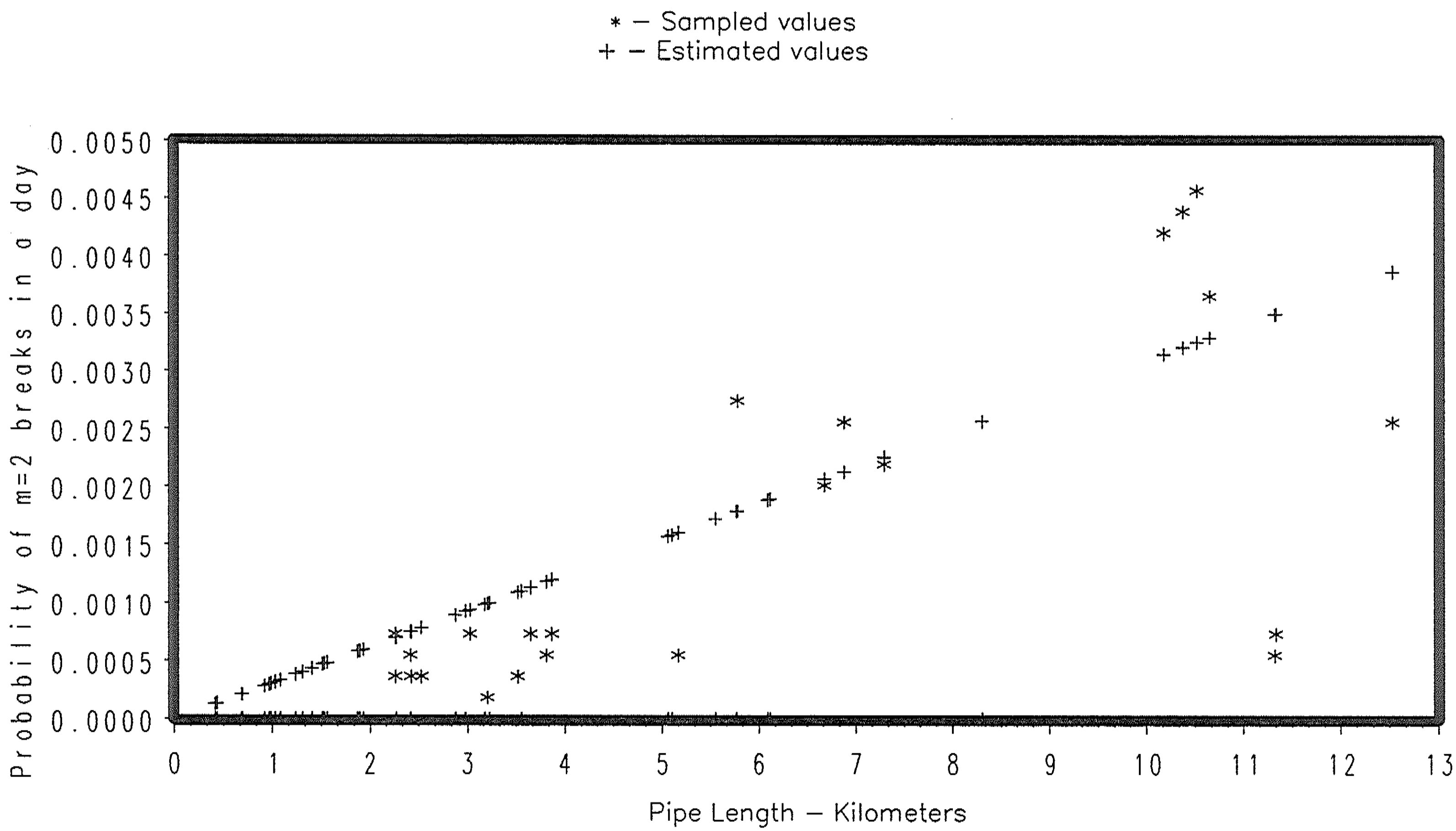


FIGURE B.68 Predicted versus sampled for m=2 breaks: cast iron pipe, 150mm 31+ years

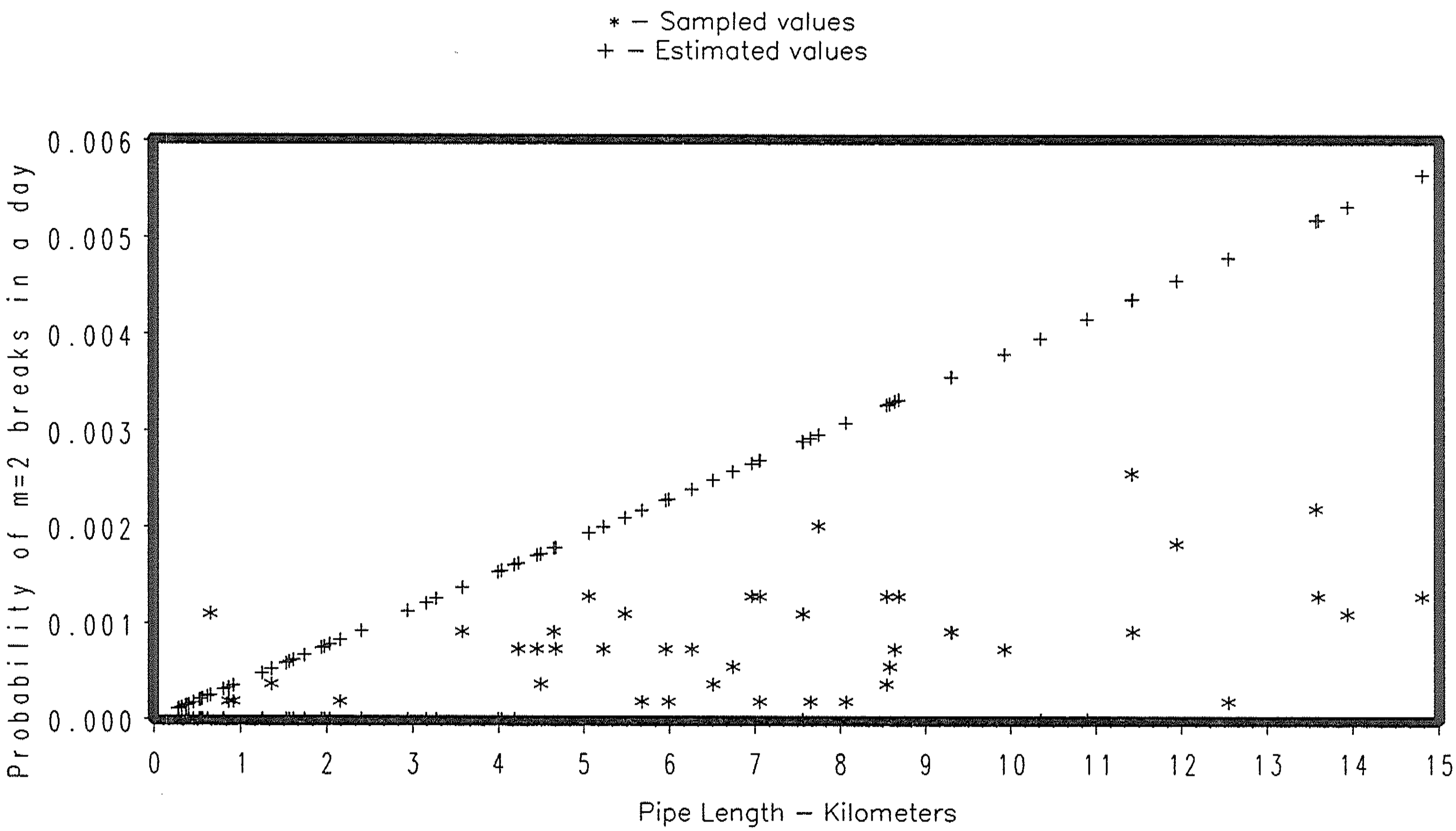


FIGURE B.69 Predicted versus sampled for m=2 breaks: cast iron pipe, 200mm 0–30 years

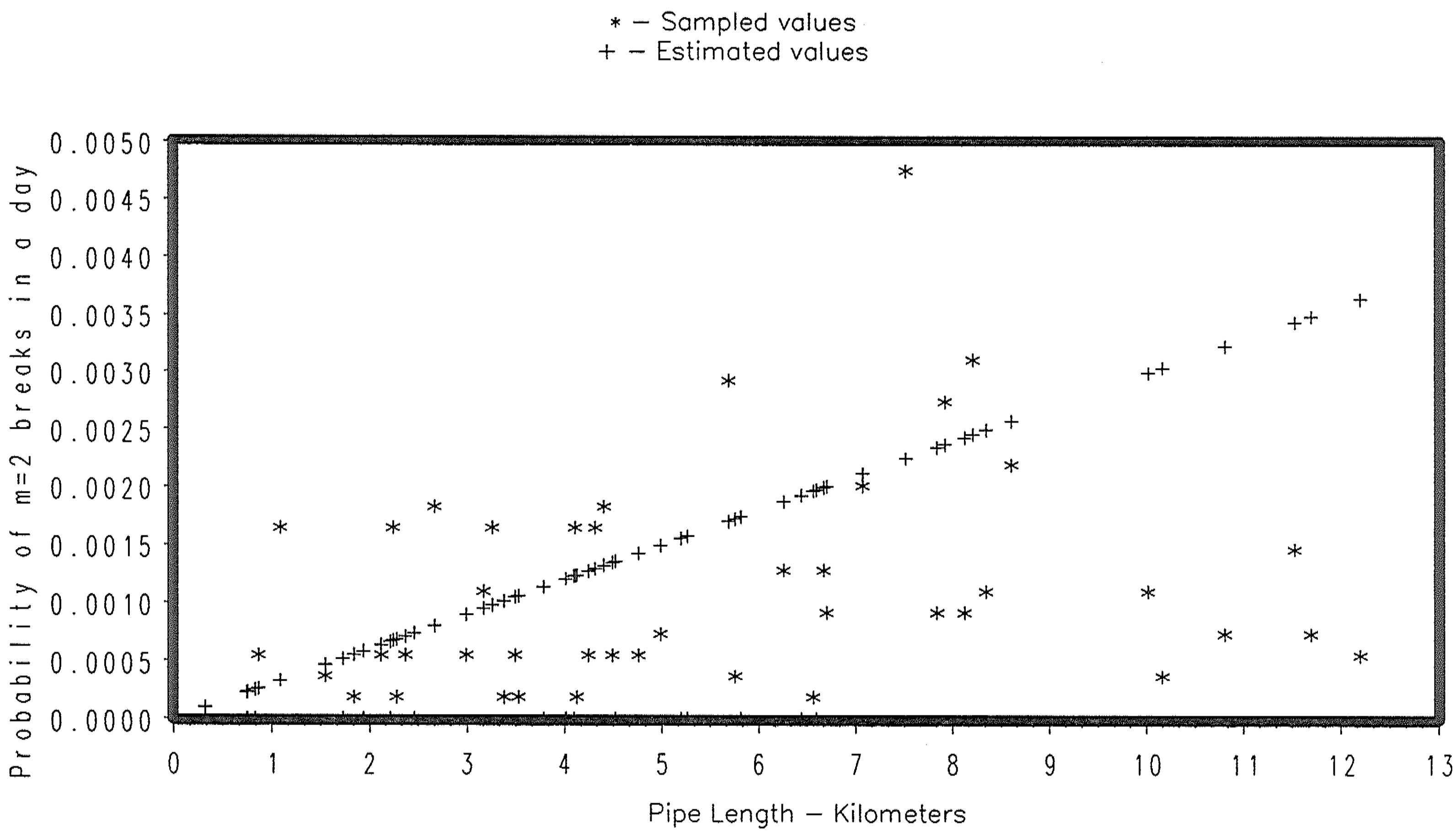


FIGURE B.70 Predicted versus sampled for m=2 breaks: cast iron pipe, 200mm 31+ years

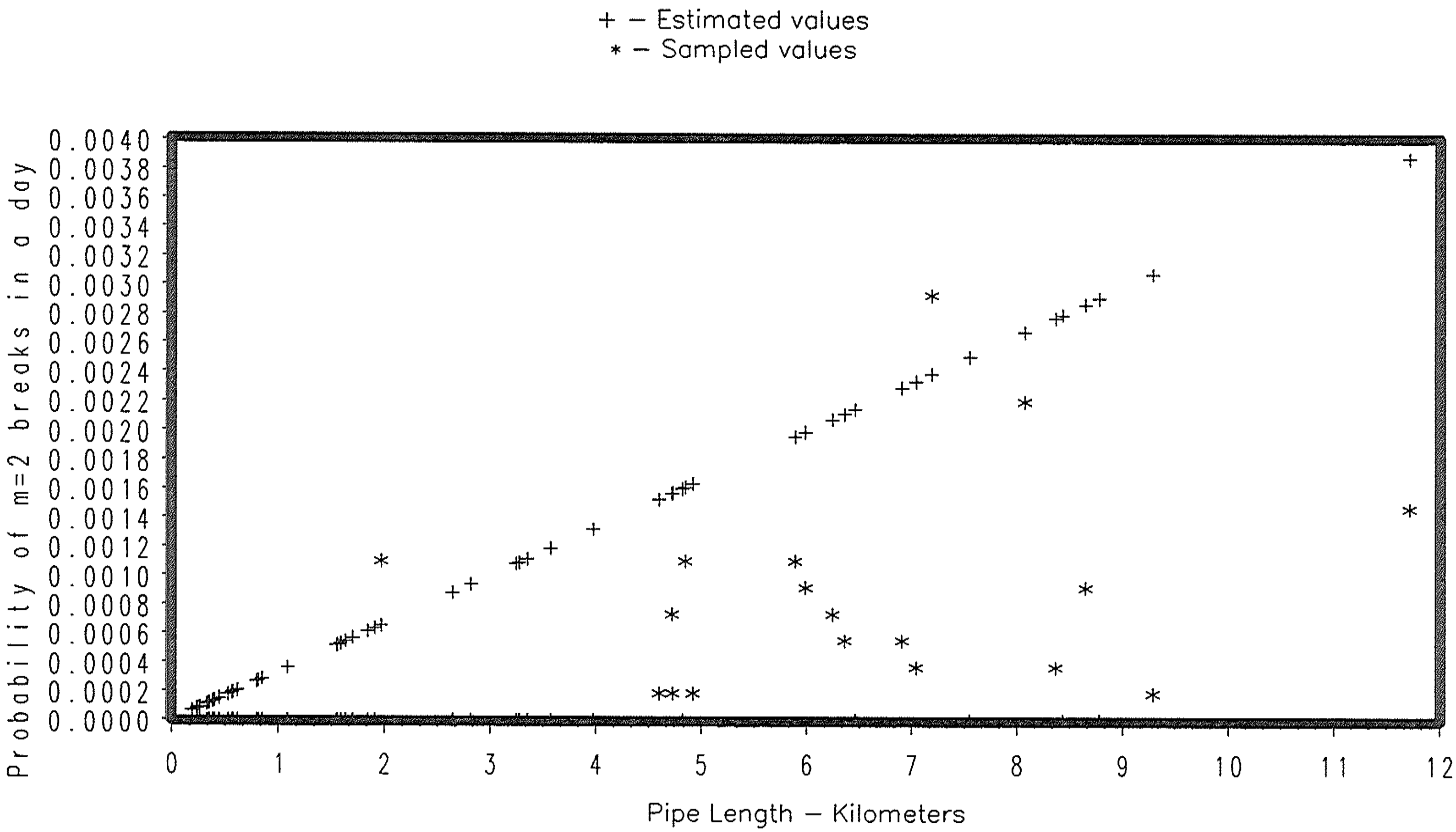


FIGURE B.71 Predicted versus sampled for m=2 breaks: cast iron pipe, 250mm 0–30 years

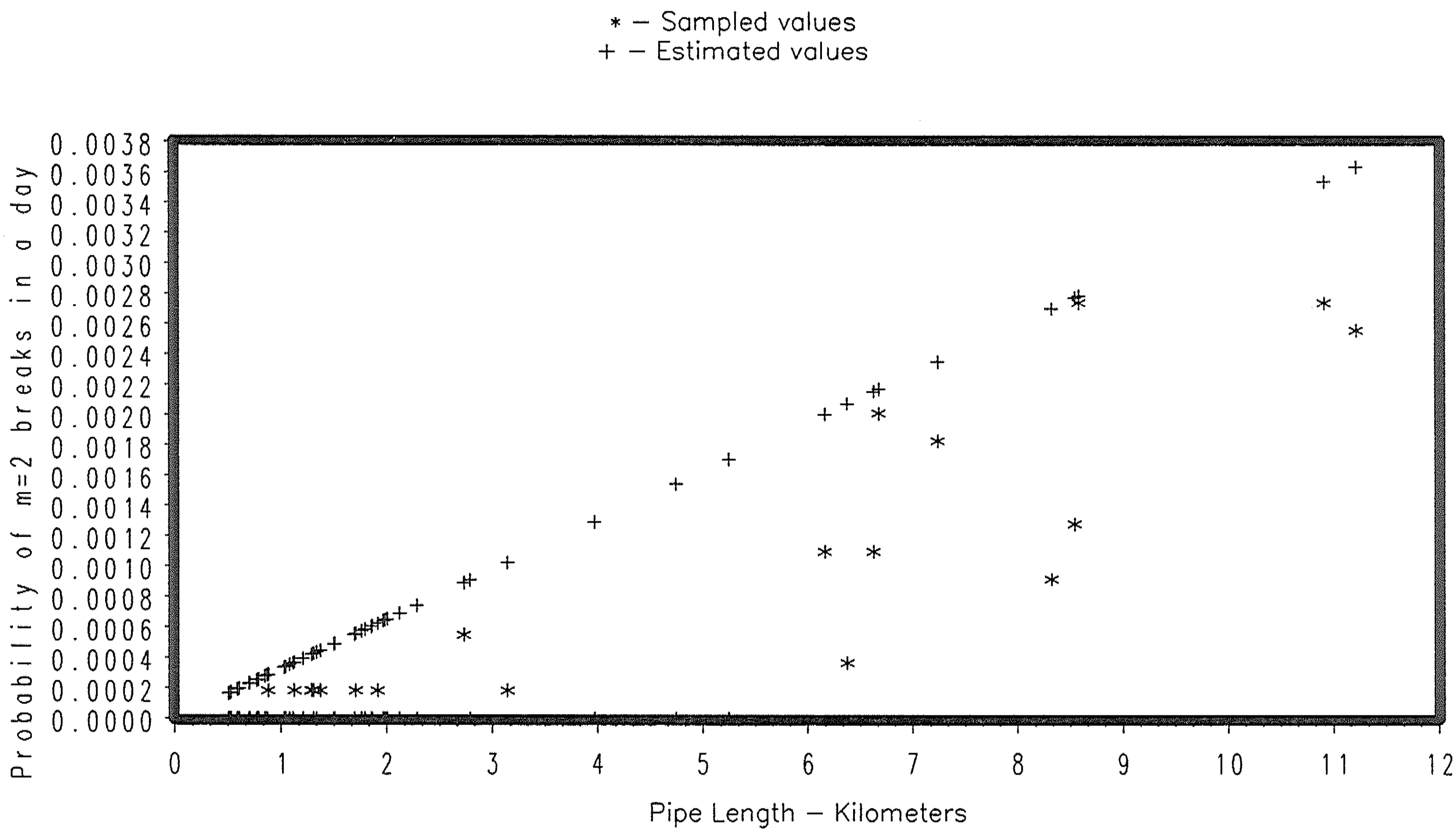


FIGURE B.72 Predicted versus sampled for m=2 breaks: cast iron pipe, 250mm 31+ years

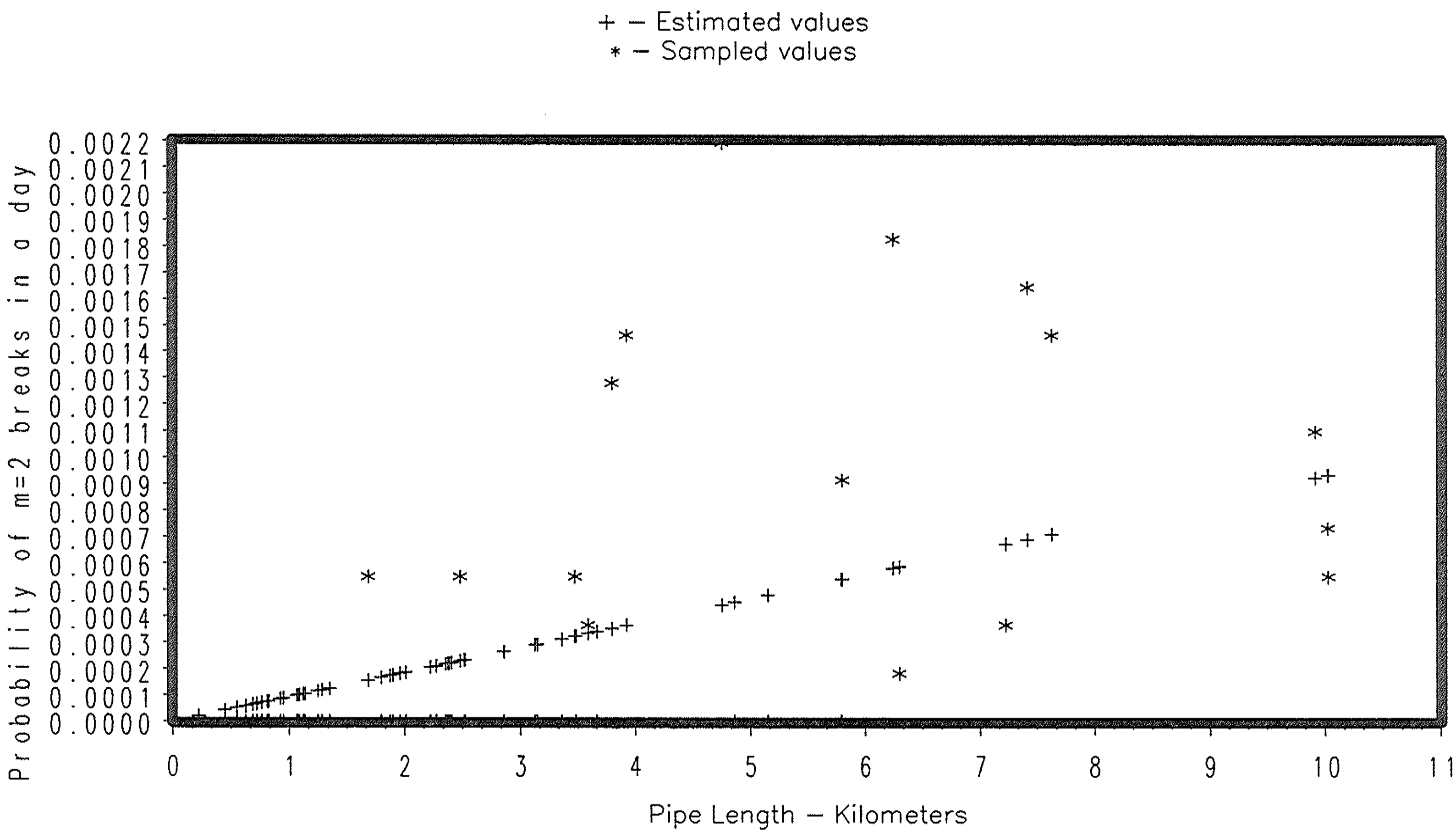


FIGURE B.73 Predicted versus sampled for m=2 breaks: cast iron pipe, 300mm 0–30 years

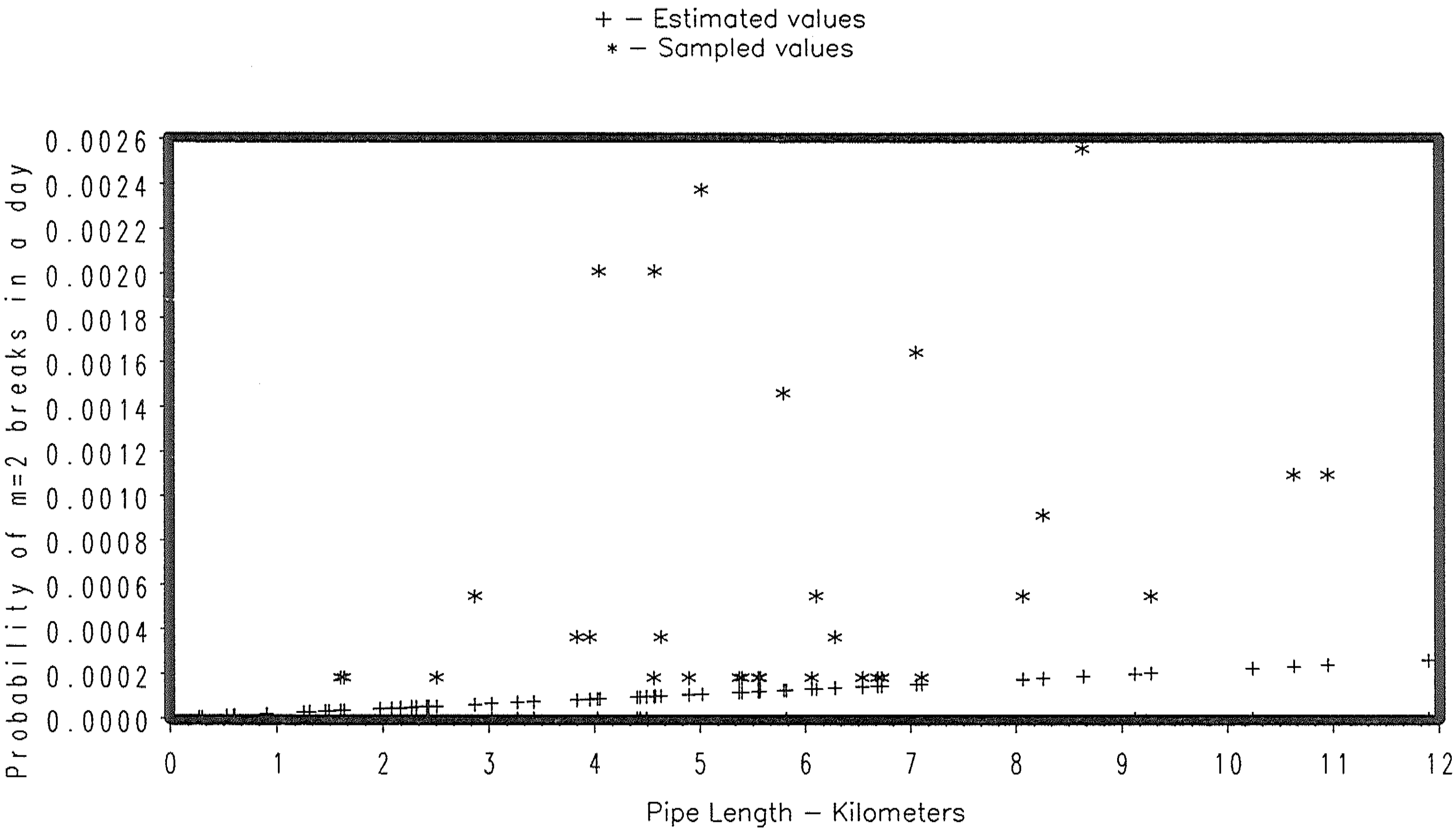


FIGURE B.74 Predicted versus sampled for m=2 breaks: cast iron pipe, 300mm 31+ years

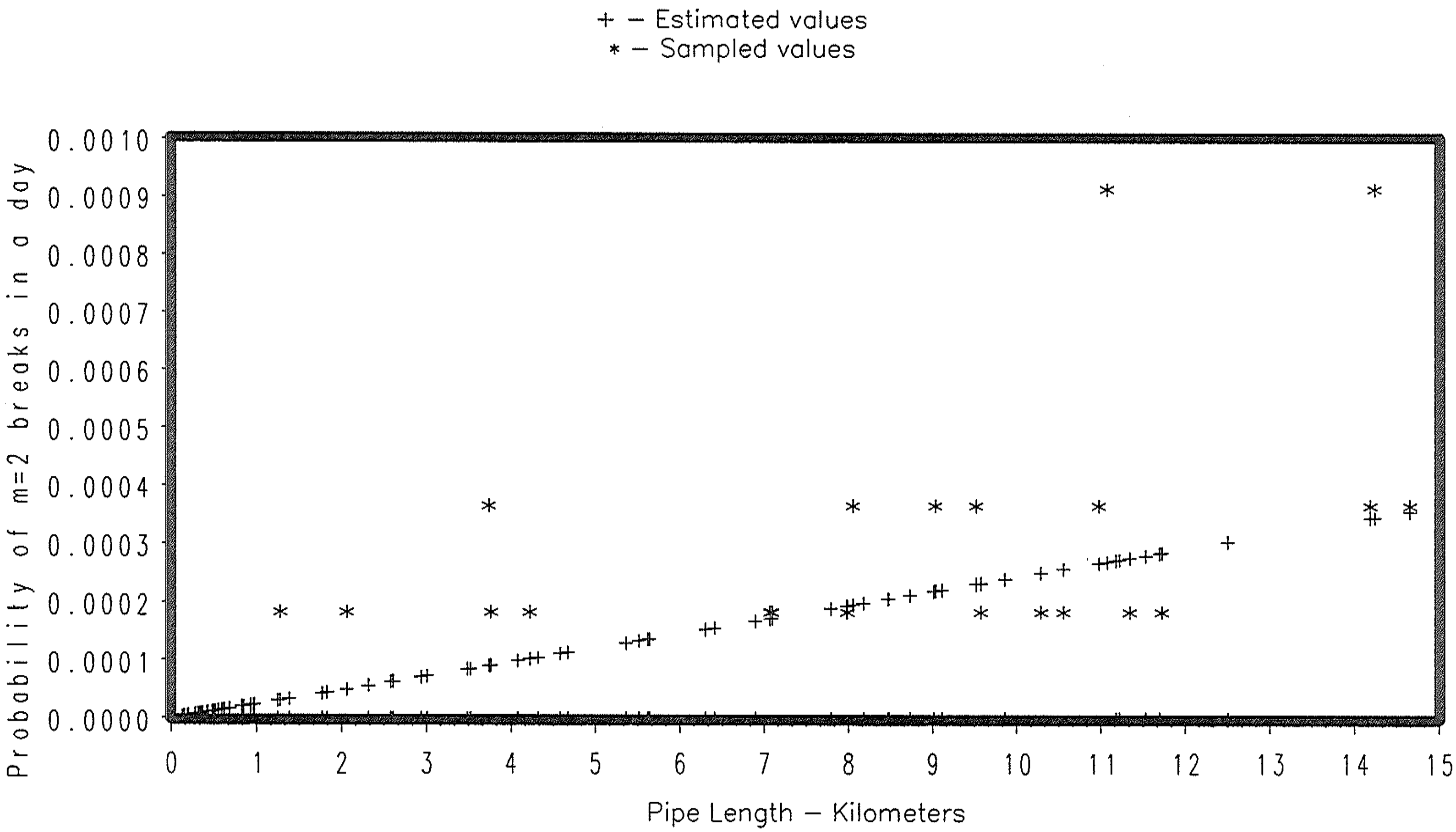


FIGURE B.75 Predicted versus sampled for m=2 breaks: asbestos cement, 150mm 0–30 years

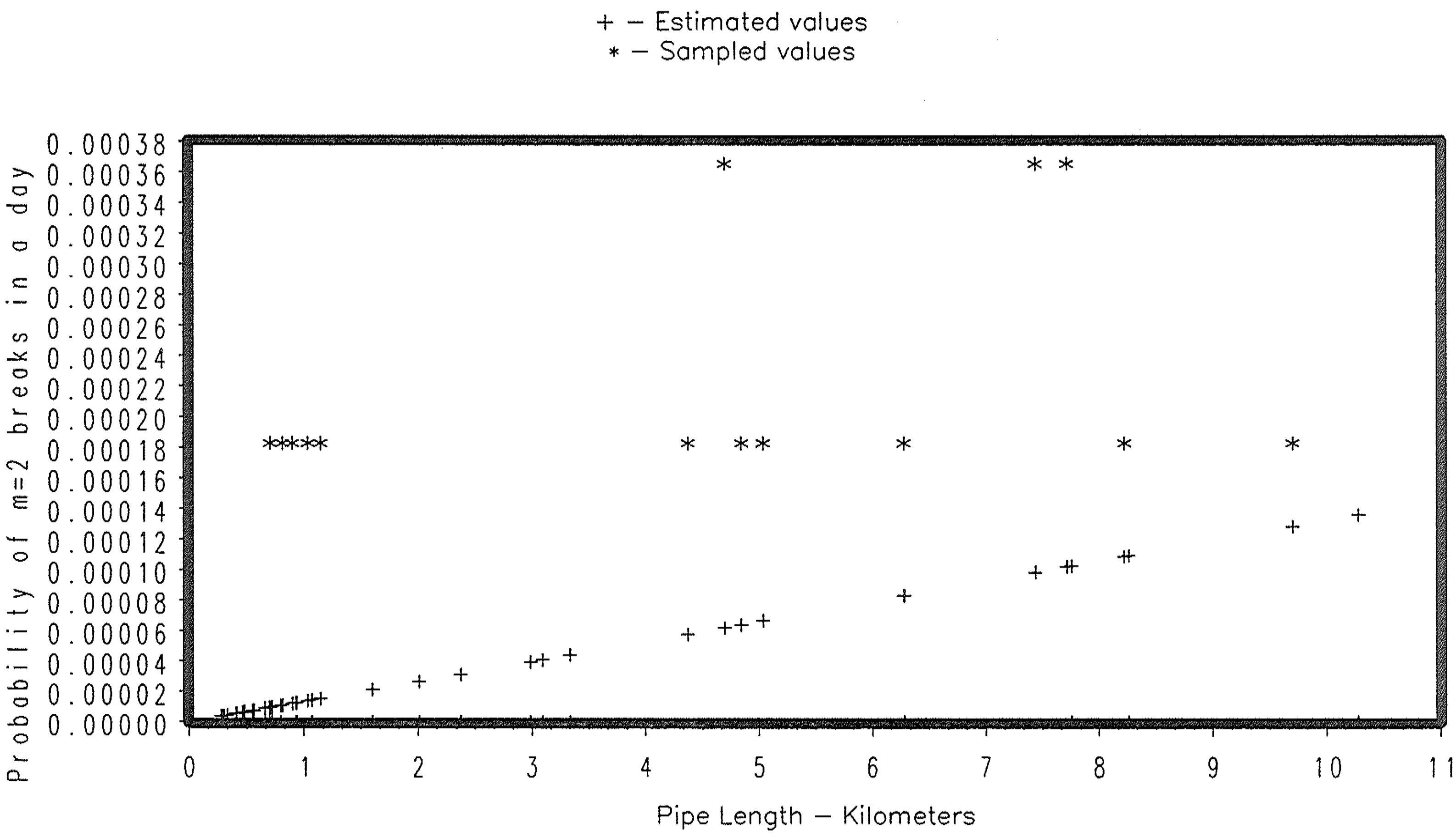


FIGURE B.76 Predicted versus sampled for m=2 breaks: asbestos cement,150mm 31+ years

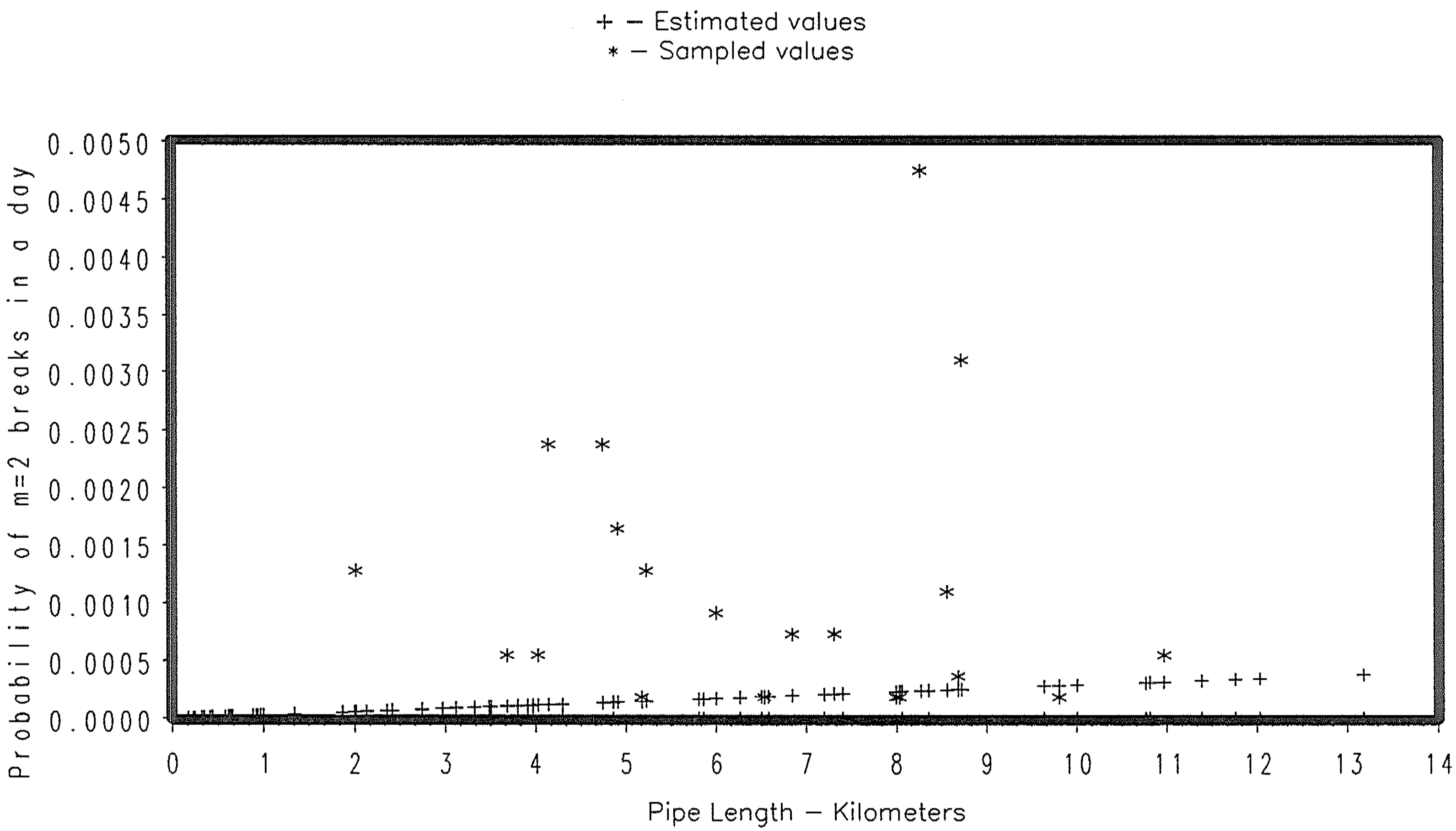


FIGURE B.77 Predicted versus sampled for m=2 breaks: asbestos cement, 200mm 0–30 years

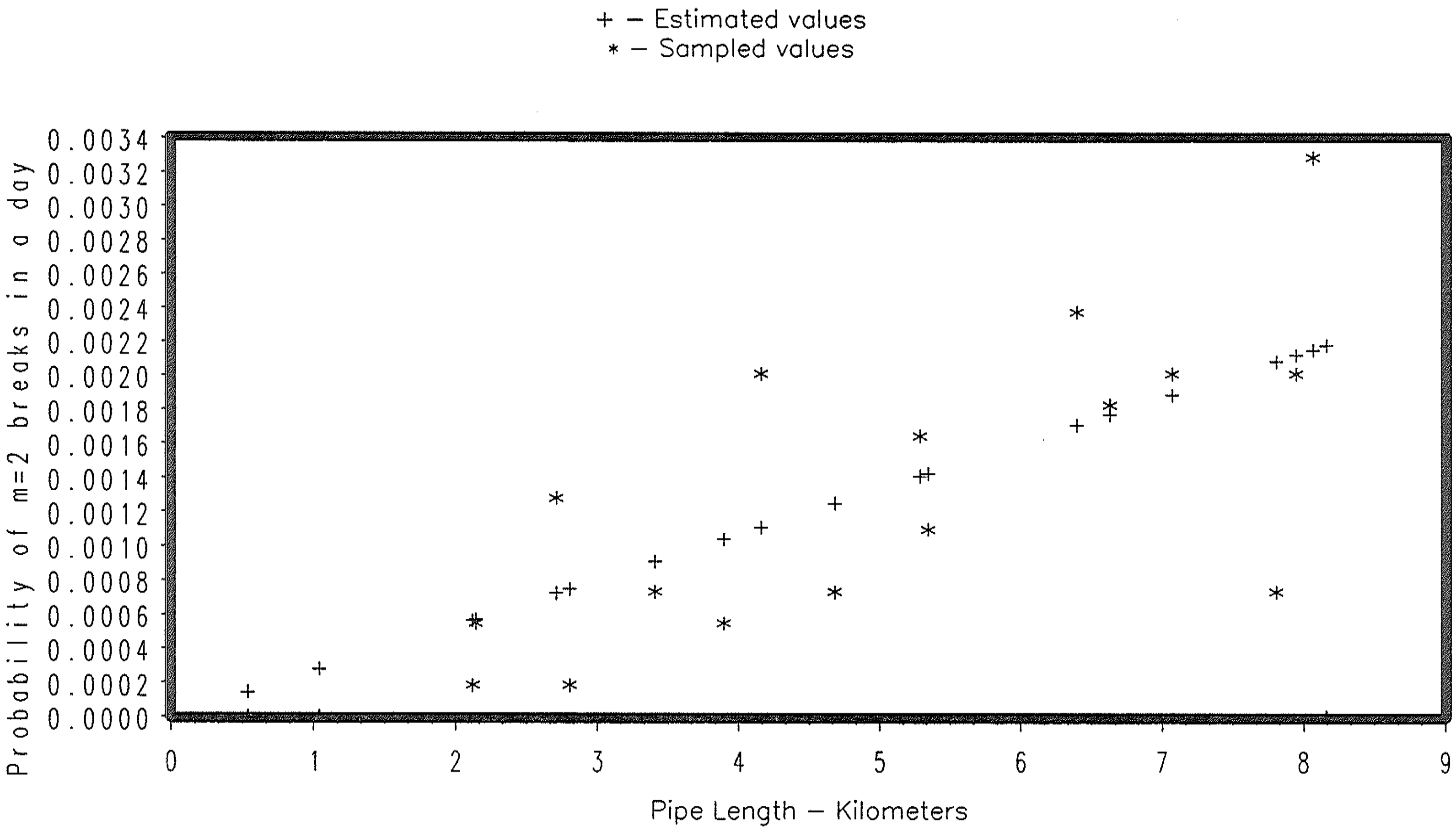


FIGURE B.78 Predicted versus sampled for m=2 breaks: asbestos cement, 200mm 31+ years

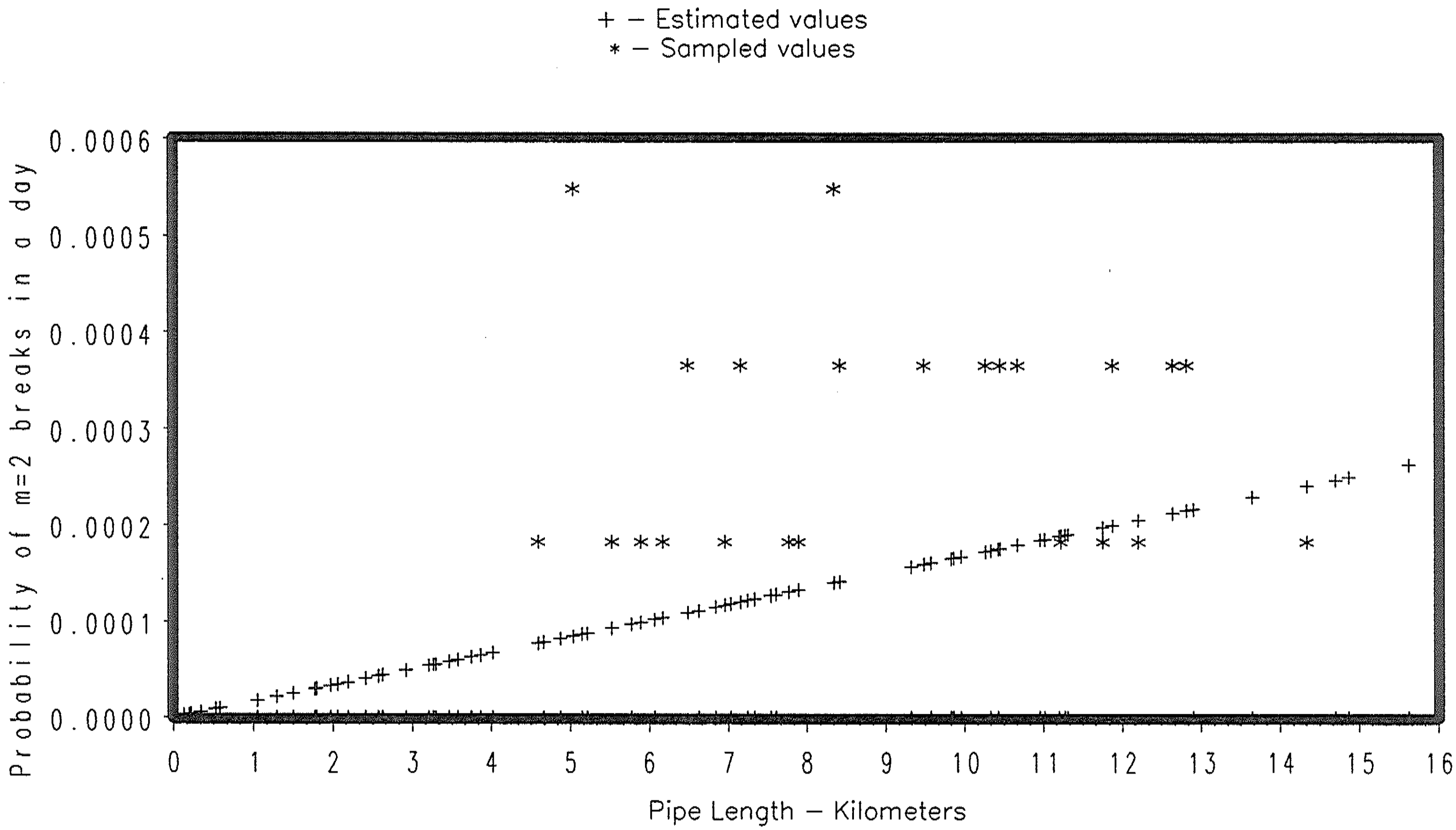


FIGURE B.79 Predicted versus sampled for m=2 breaks: asbestos cement, 250mm 0+ years

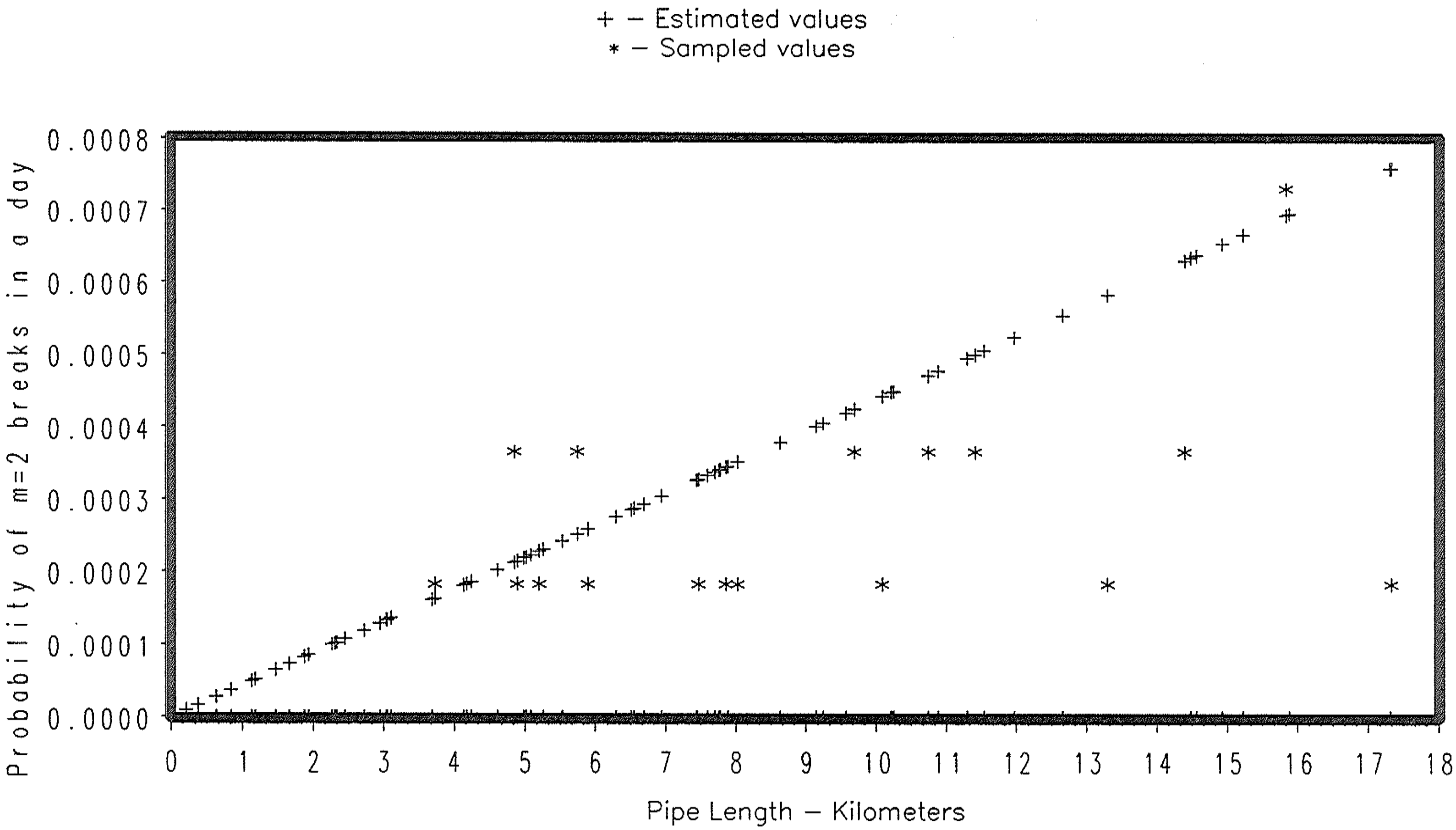


FIGURE B.80 Predicted versus sampled for m=2 breaks: asbestos cement, 300mm 0+ years

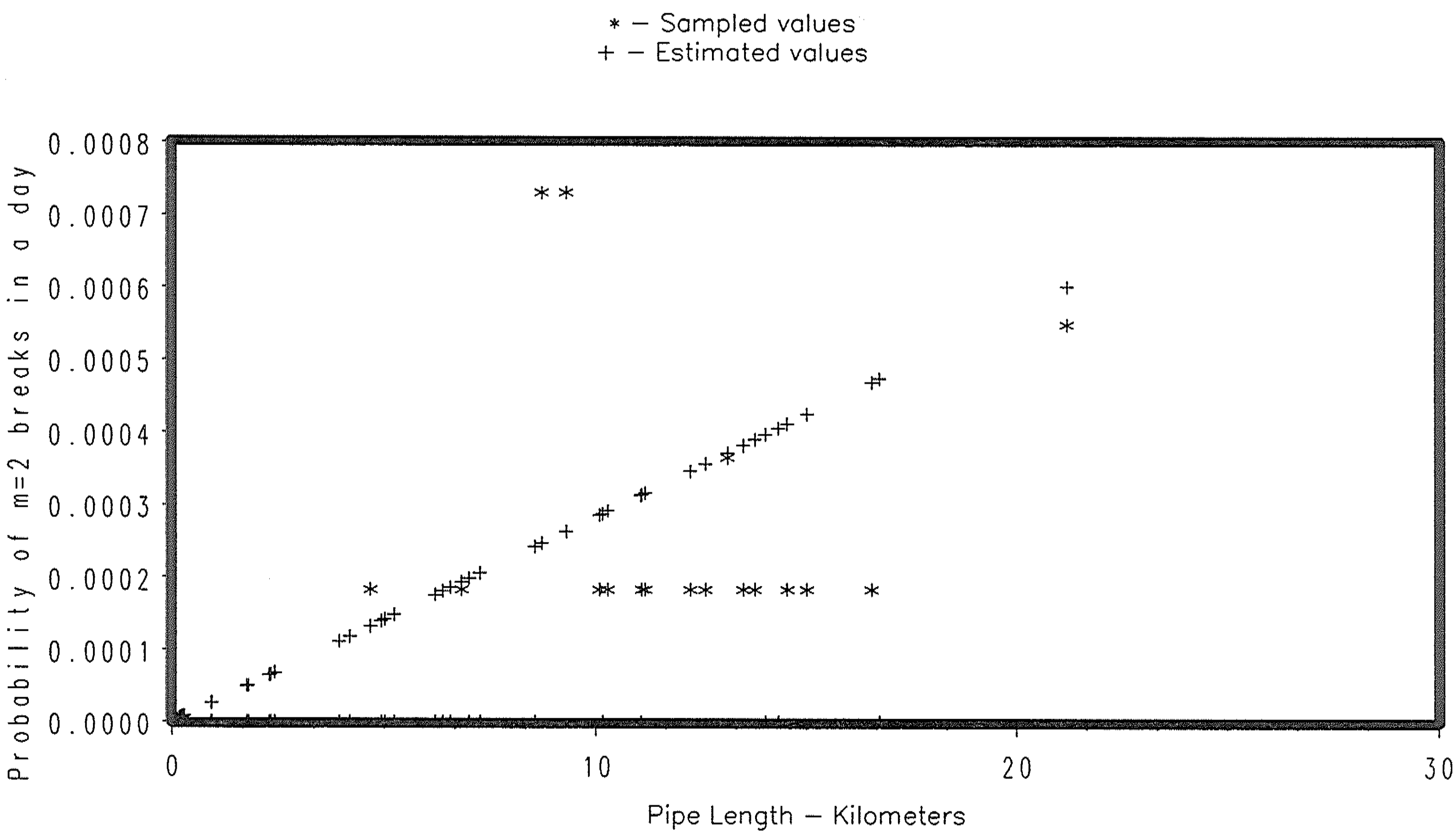


FIGURE B.81 Predicted versus sampled for m=2 breaks: ductile steel, 150mm 0+ years

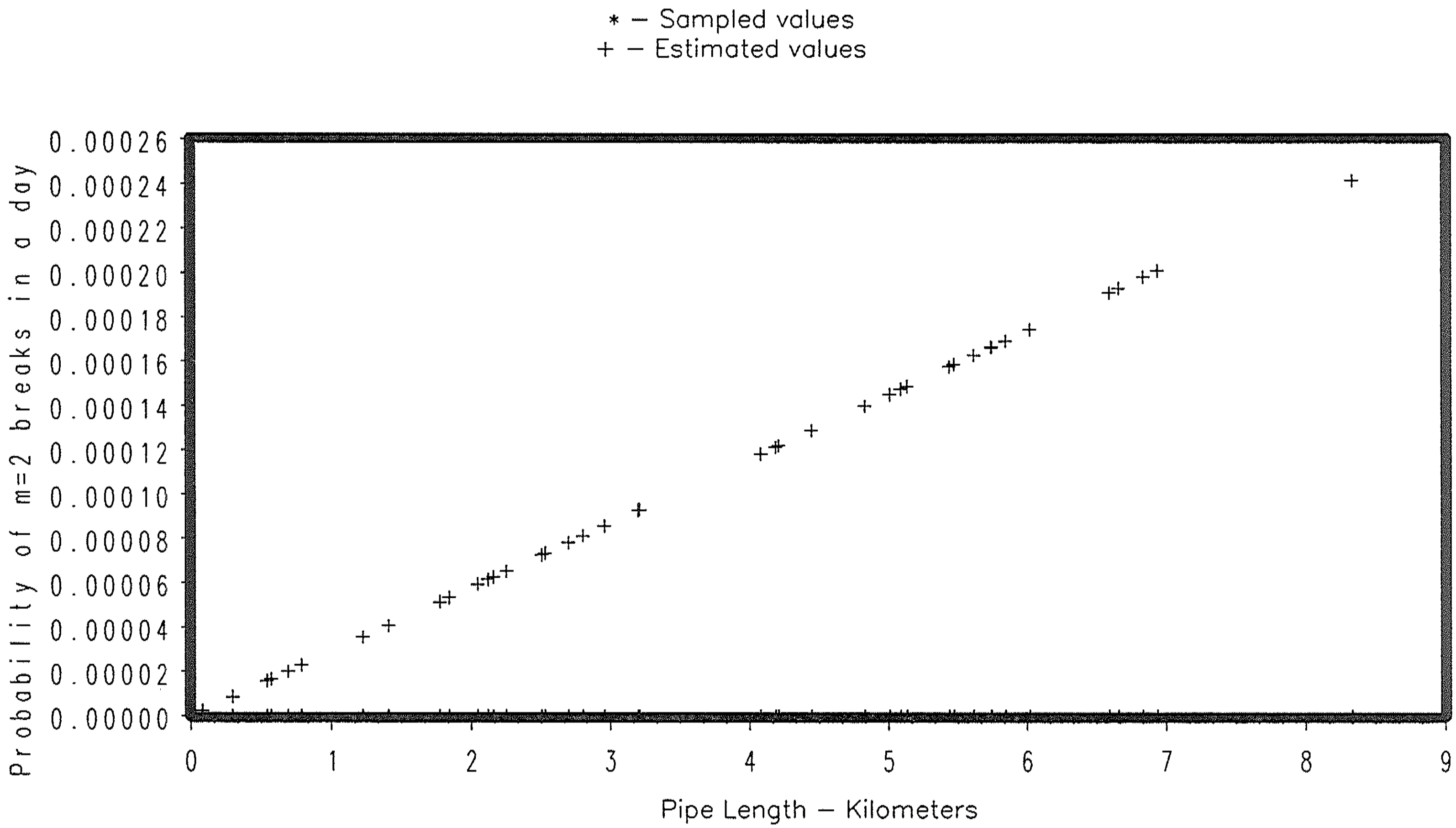


FIGURE B.82 Predicted versus sampled for m=2 breaks: ductile steel, 200mm 0+ years

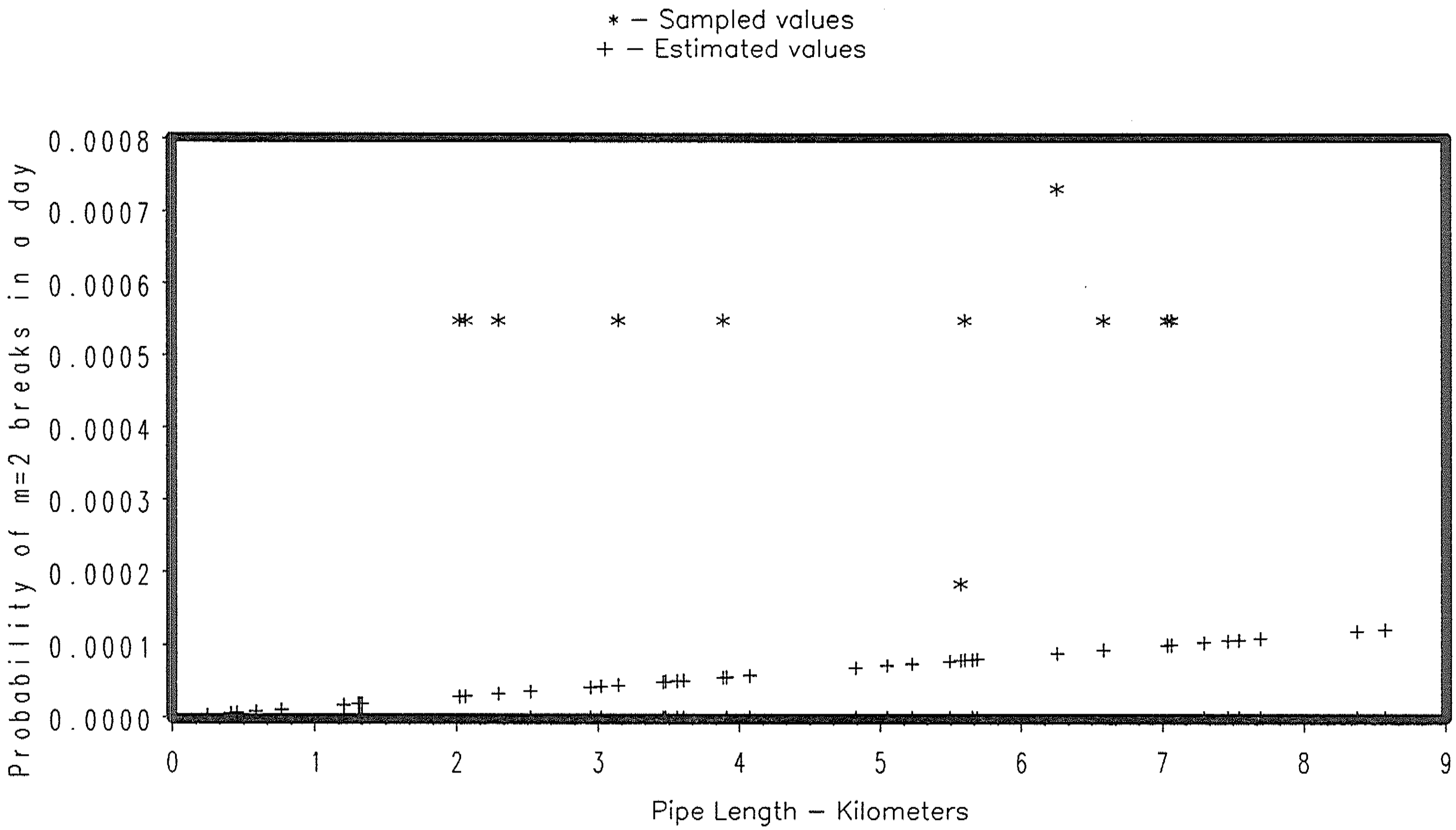


FIGURE B.83 Predicted versus sampled for m=2 breaks: ductile steel, 250mm 0+ years

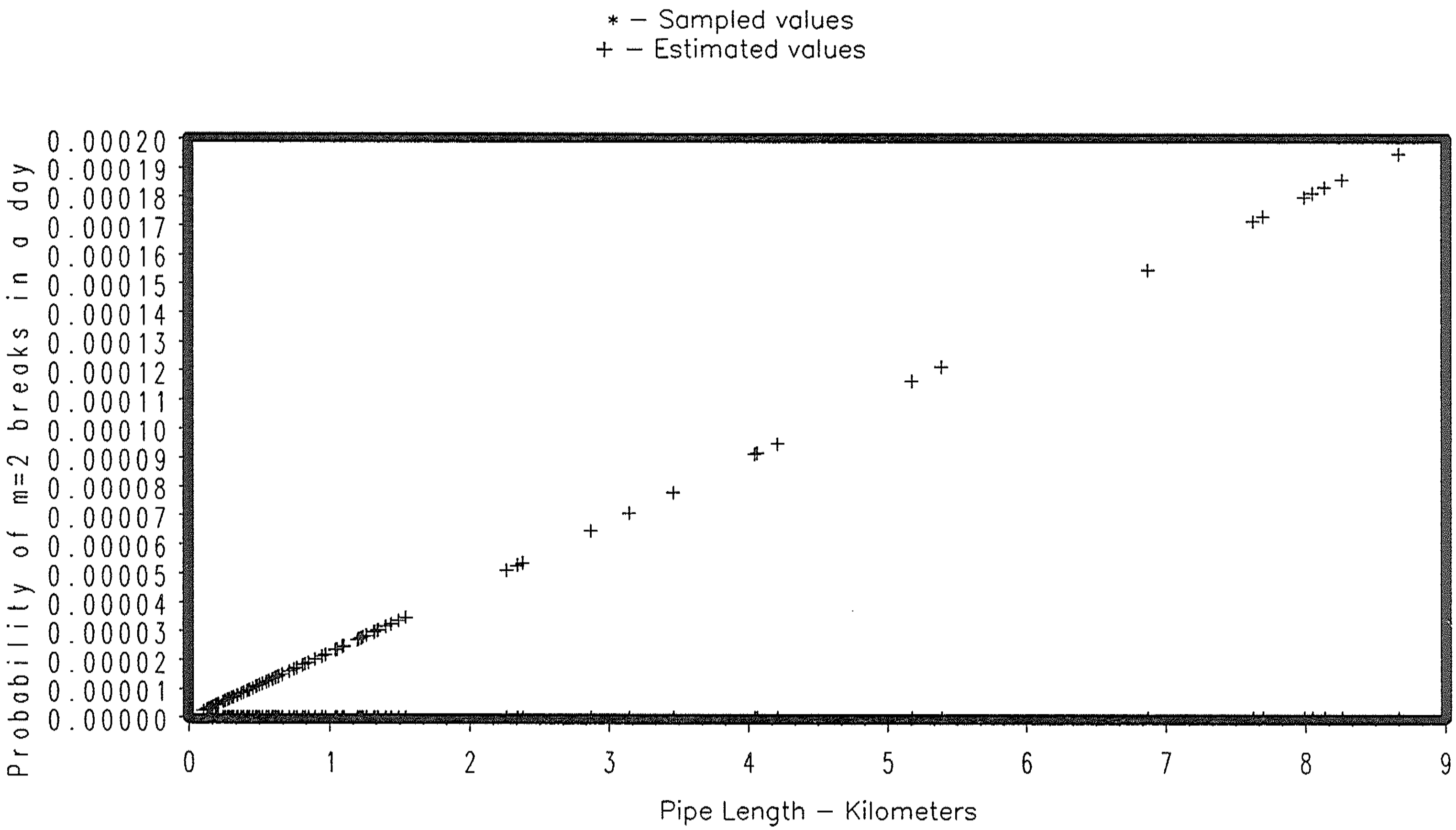


FIGURE B.84 Predicted versus sampled for m=2 breaks: PVC, 150mm 0+ years

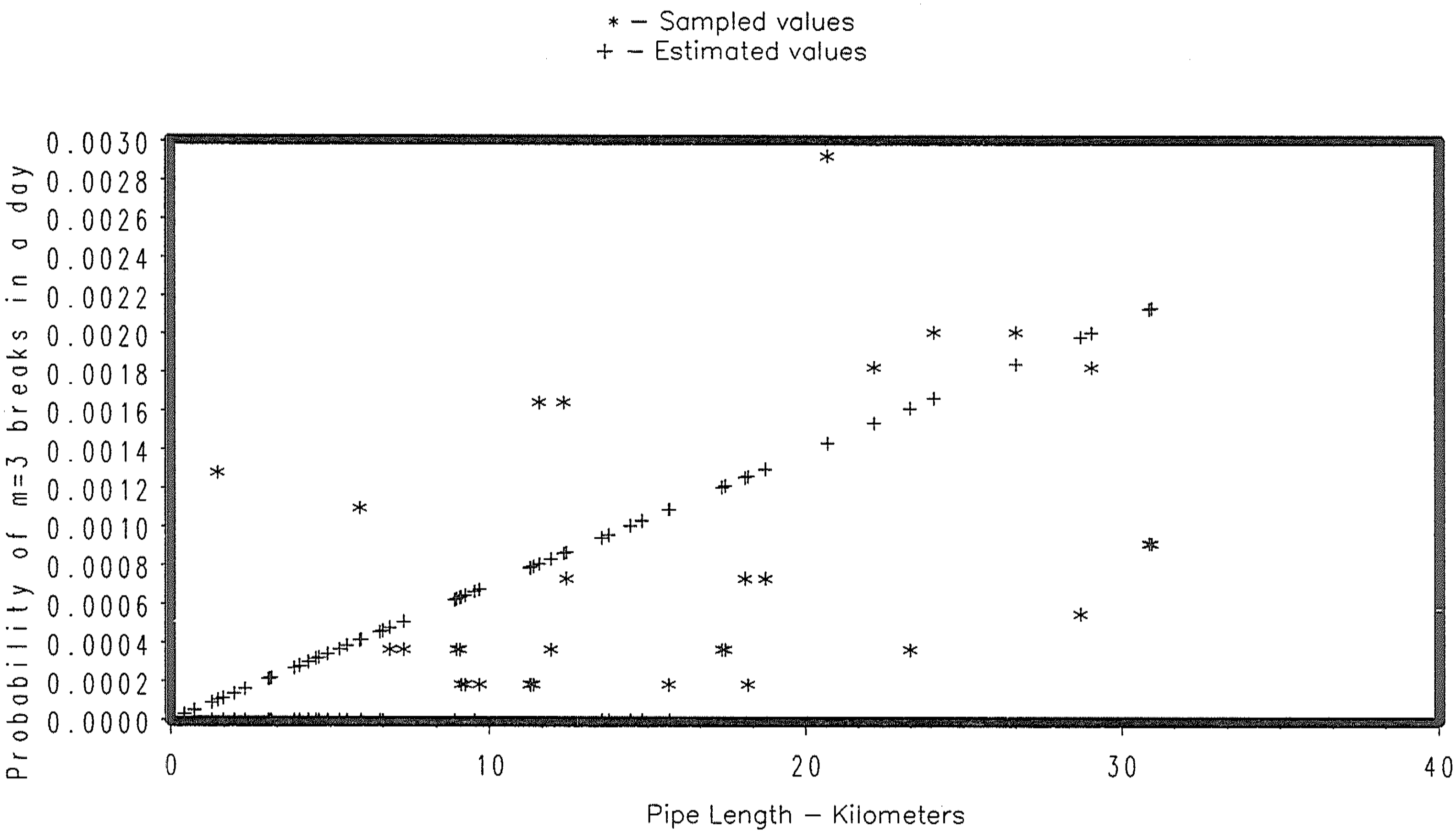


FIGURE B.85 Predicted versus sampled for m=3 breaks: cast iron pipe, 150mm 0–18 years

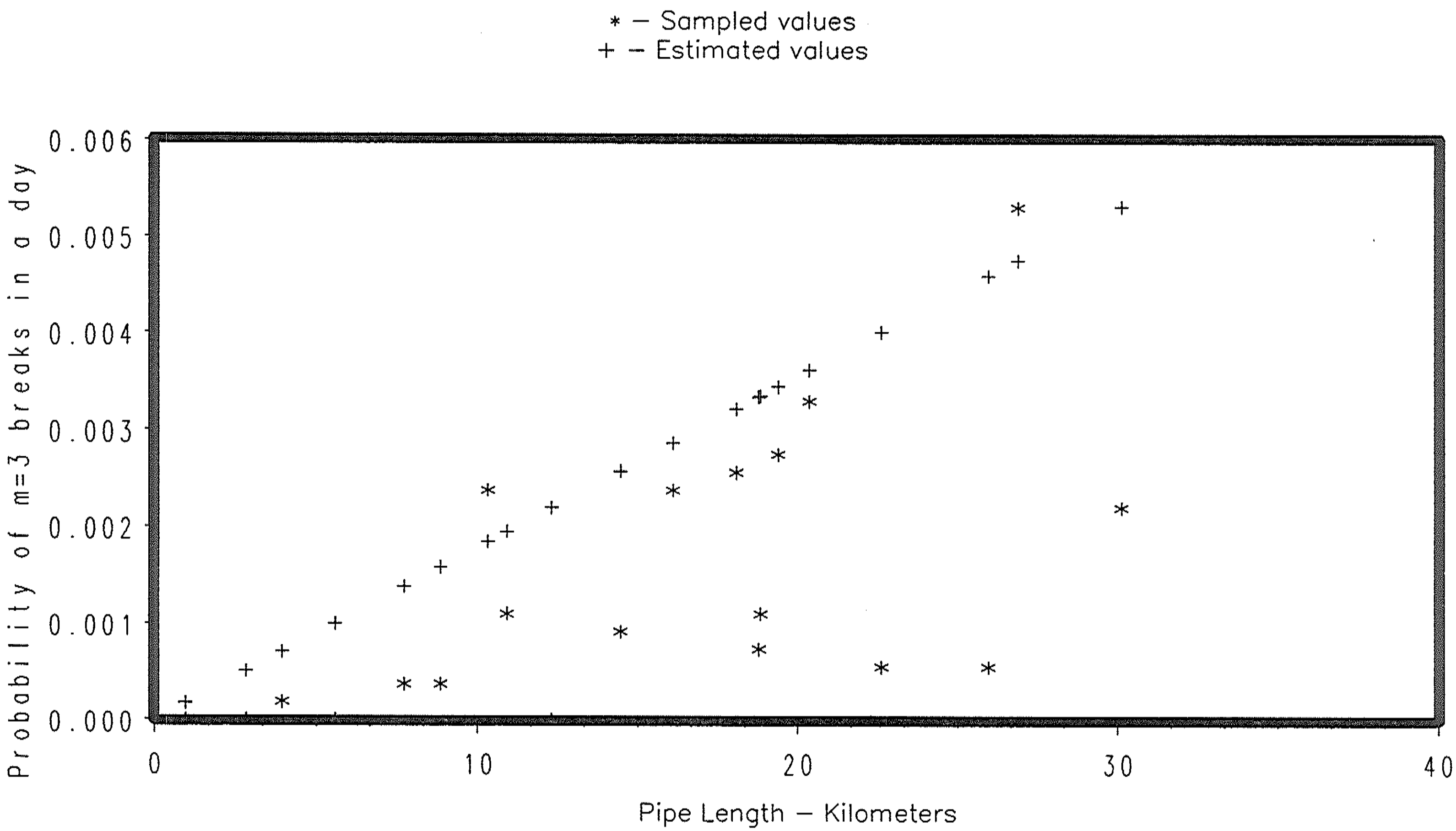


FIGURE B.86 Predicted versus sampled for m=3 breaks: cast iron pipe, 150mm 19–30 years

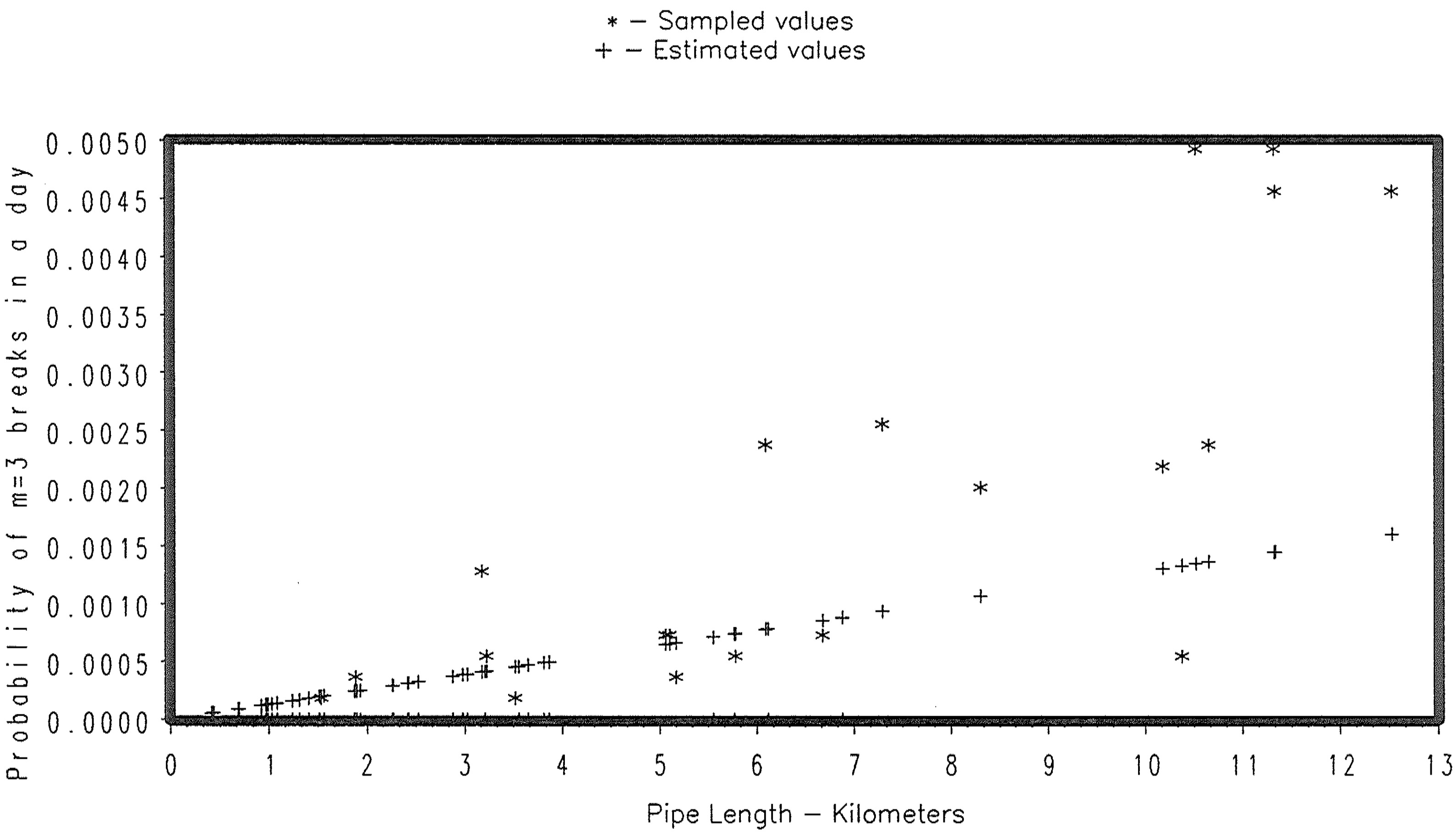


FIGURE B.87 Predicted versus sampled for m=3 breaks: cast iron pipe, 150mm 31+ years

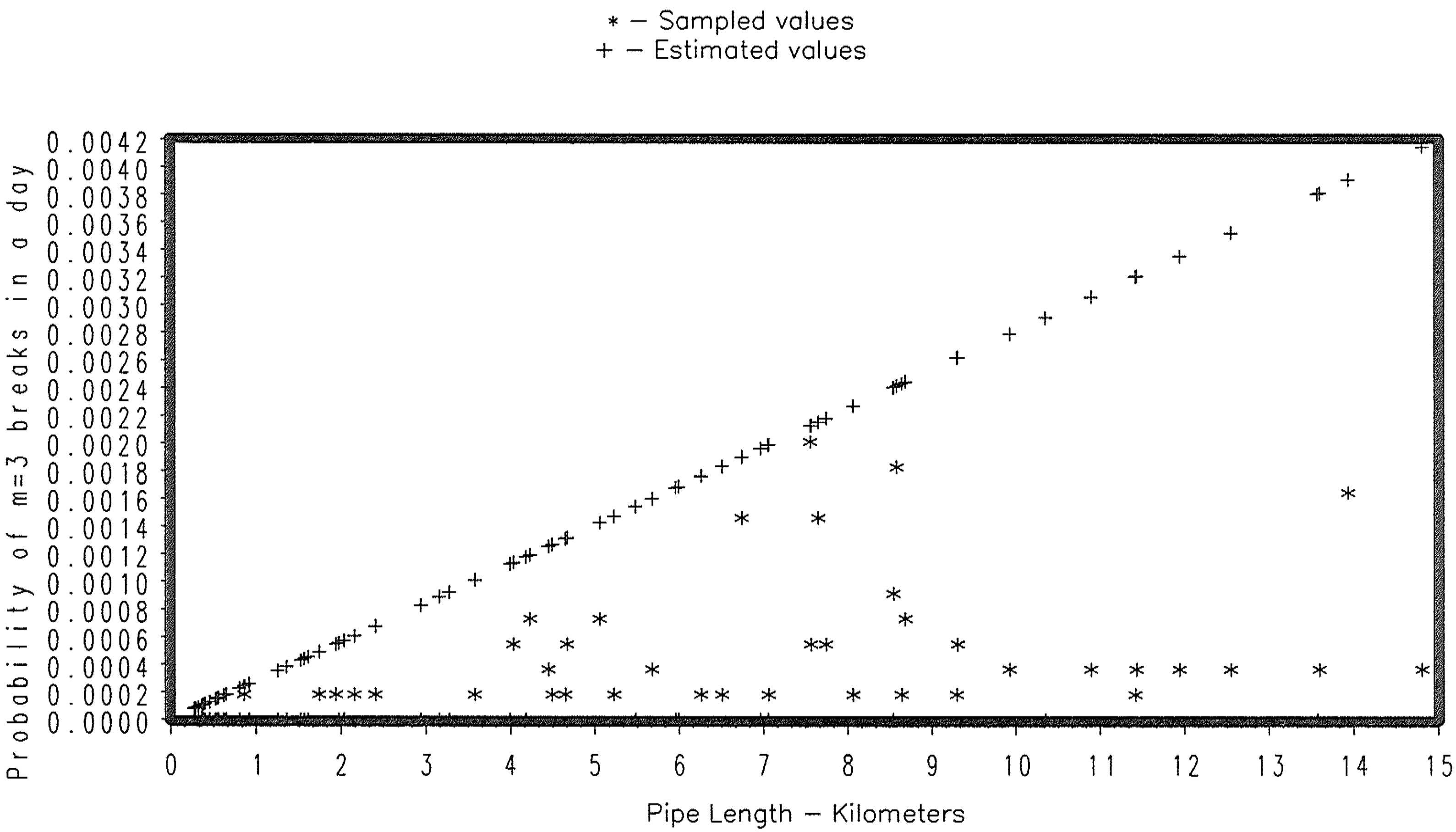


FIGURE B.88 Predicted versus sampled for m=3 breaks: cast iron pipe, 200mm 0–30 years

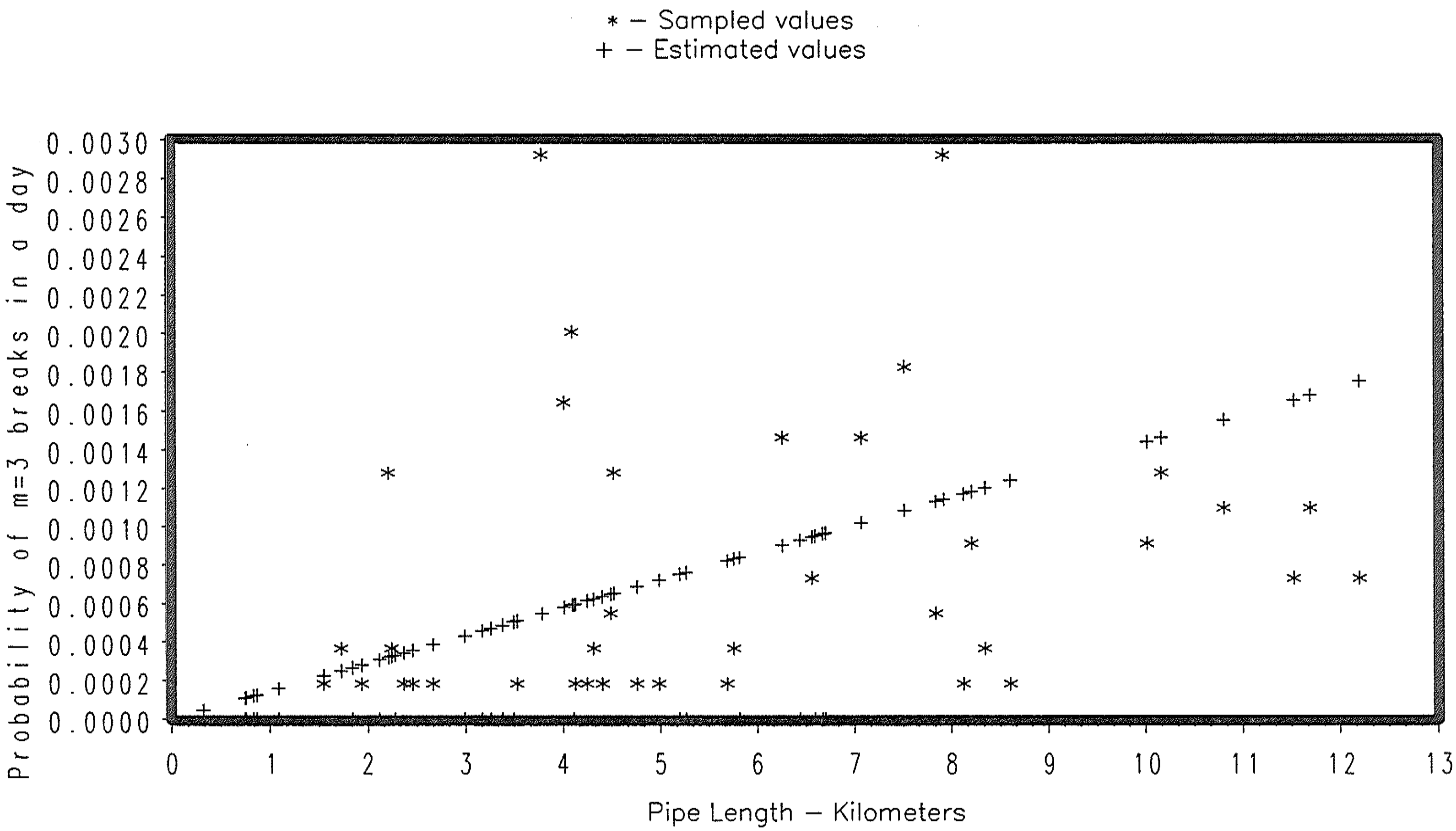


FIGURE B.89 Predicted versus sampled for m=3 breaks: cast iron pipe, 200mm 31+ years

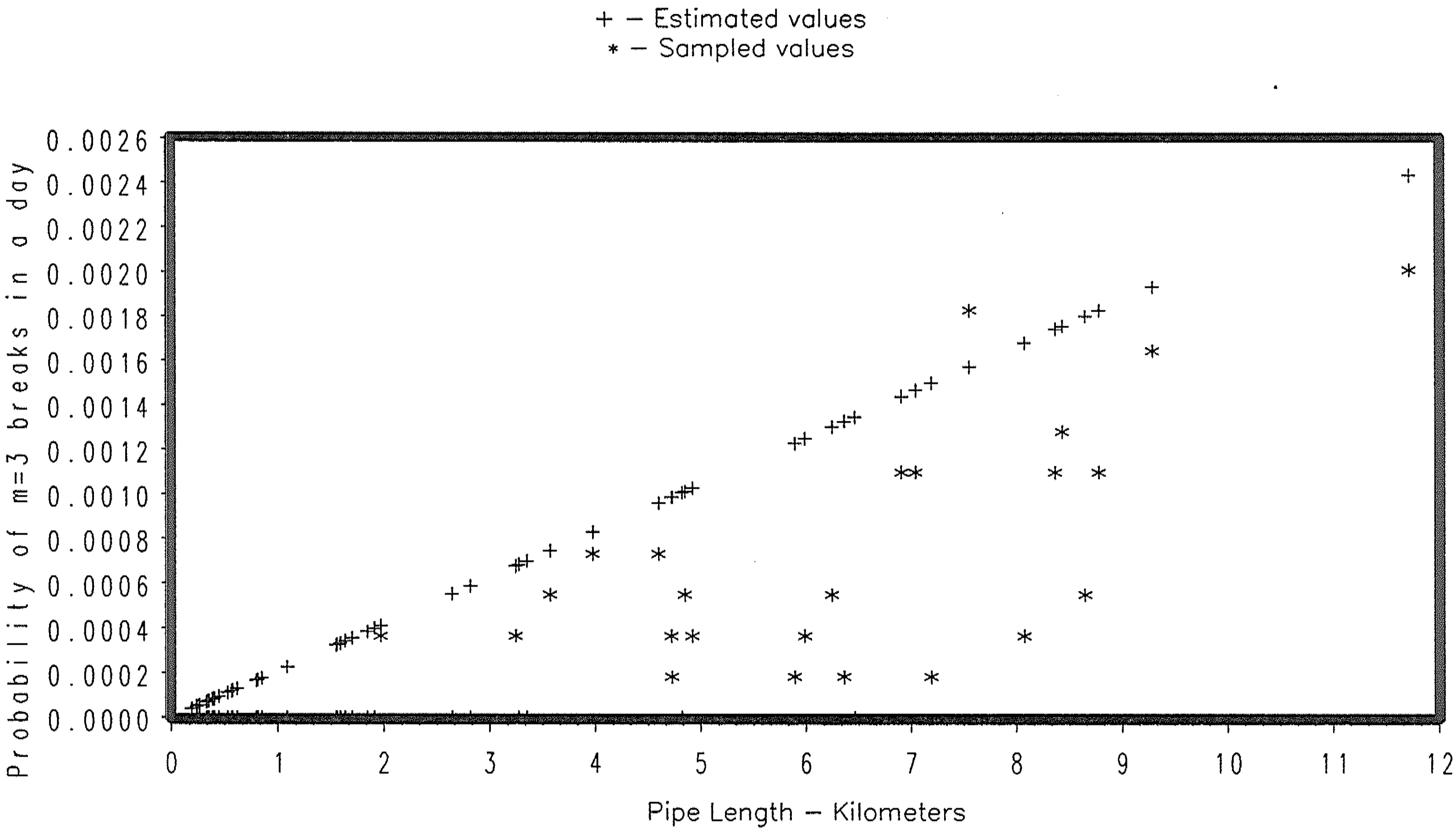


FIGURE B.90 Predicted versus sampled for m=3 breaks: cast iron pipe, 250mm 0–30 years

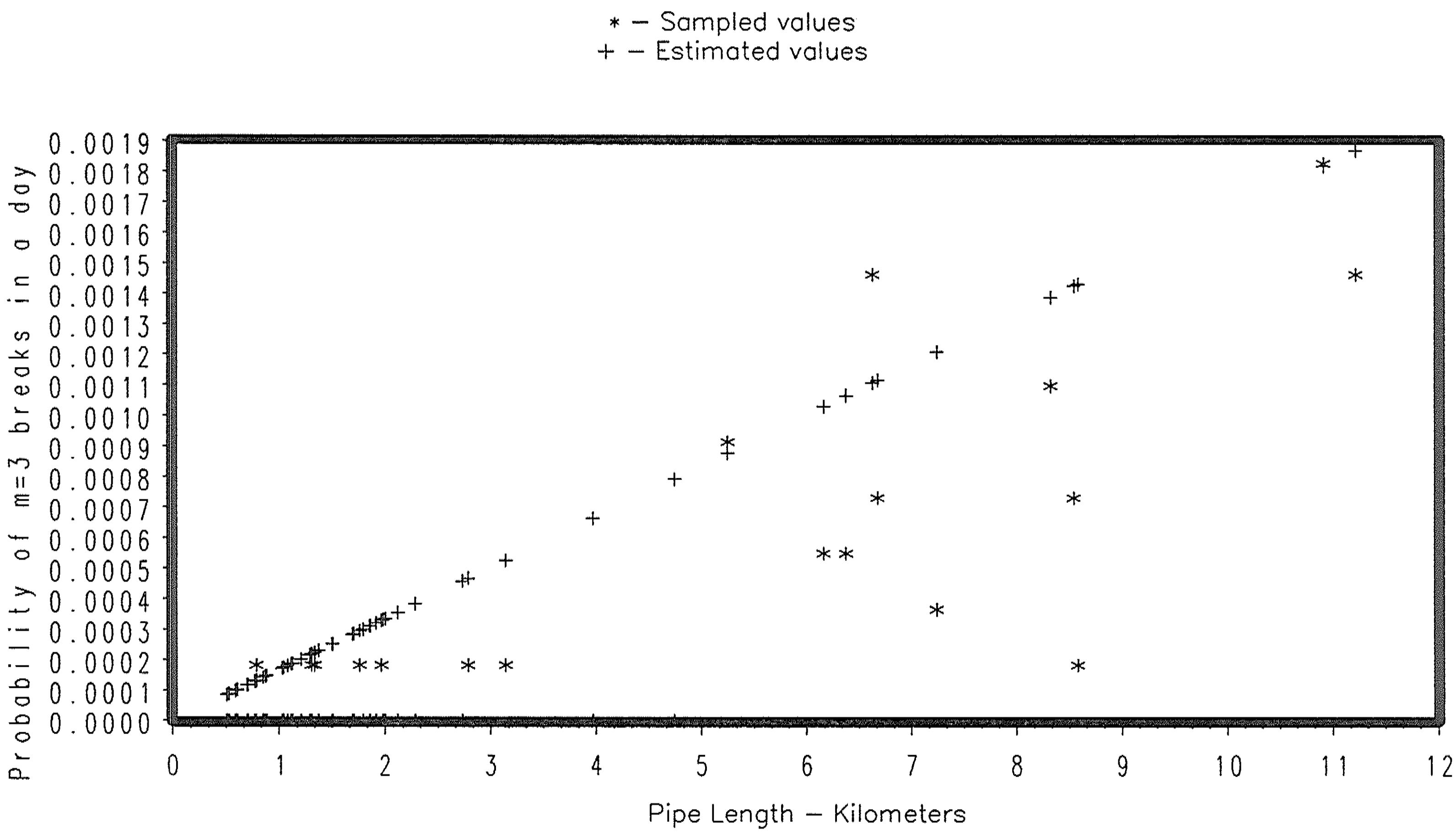


FIGURE B.91 Predicted versus sampled for m=3 breaks: cast iron pipe, 250mm 31+ years

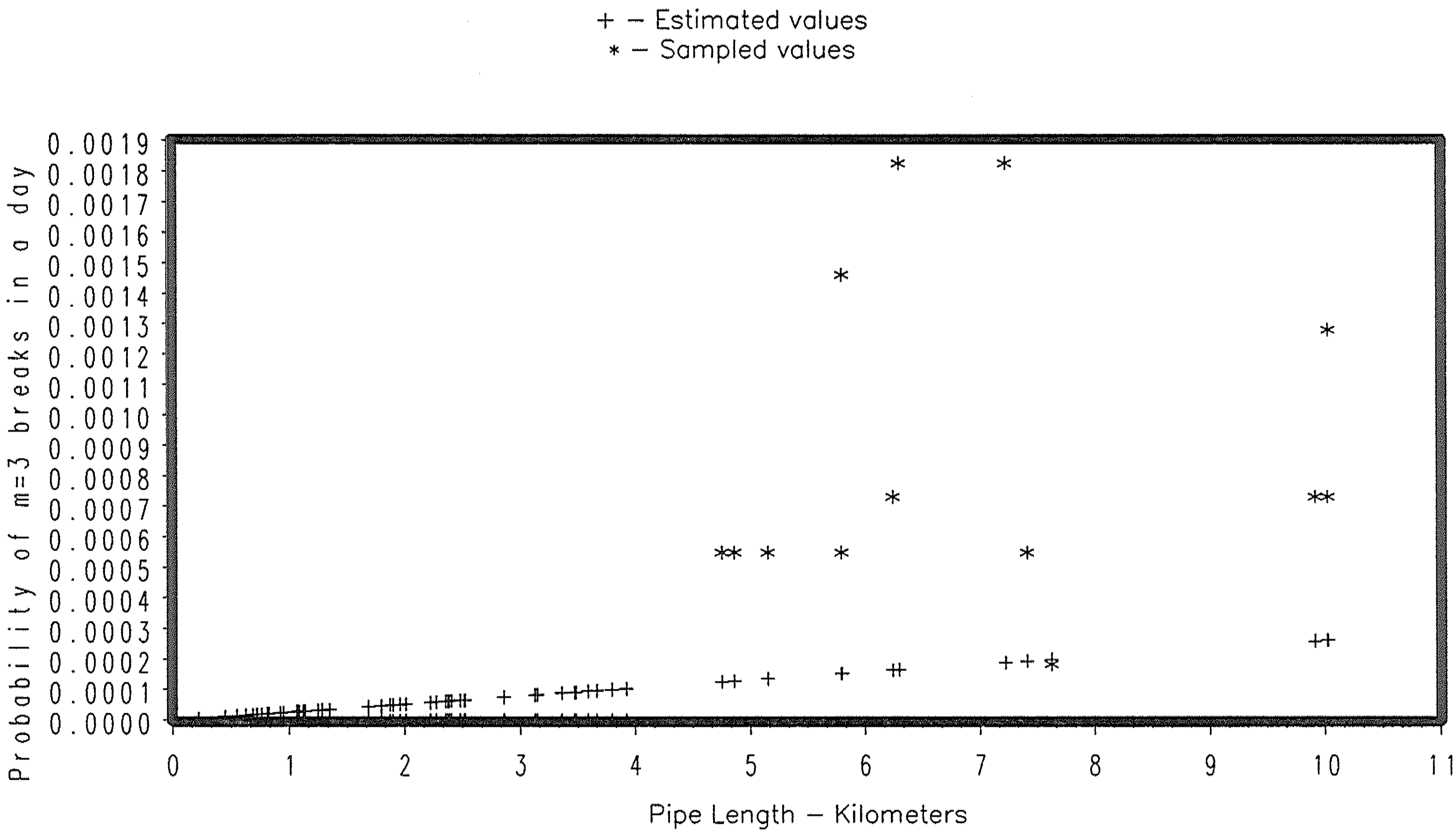


FIGURE B.92 Predicted versus sampled for m=3 breaks: cast iron pipe, 300mm 0–30 years

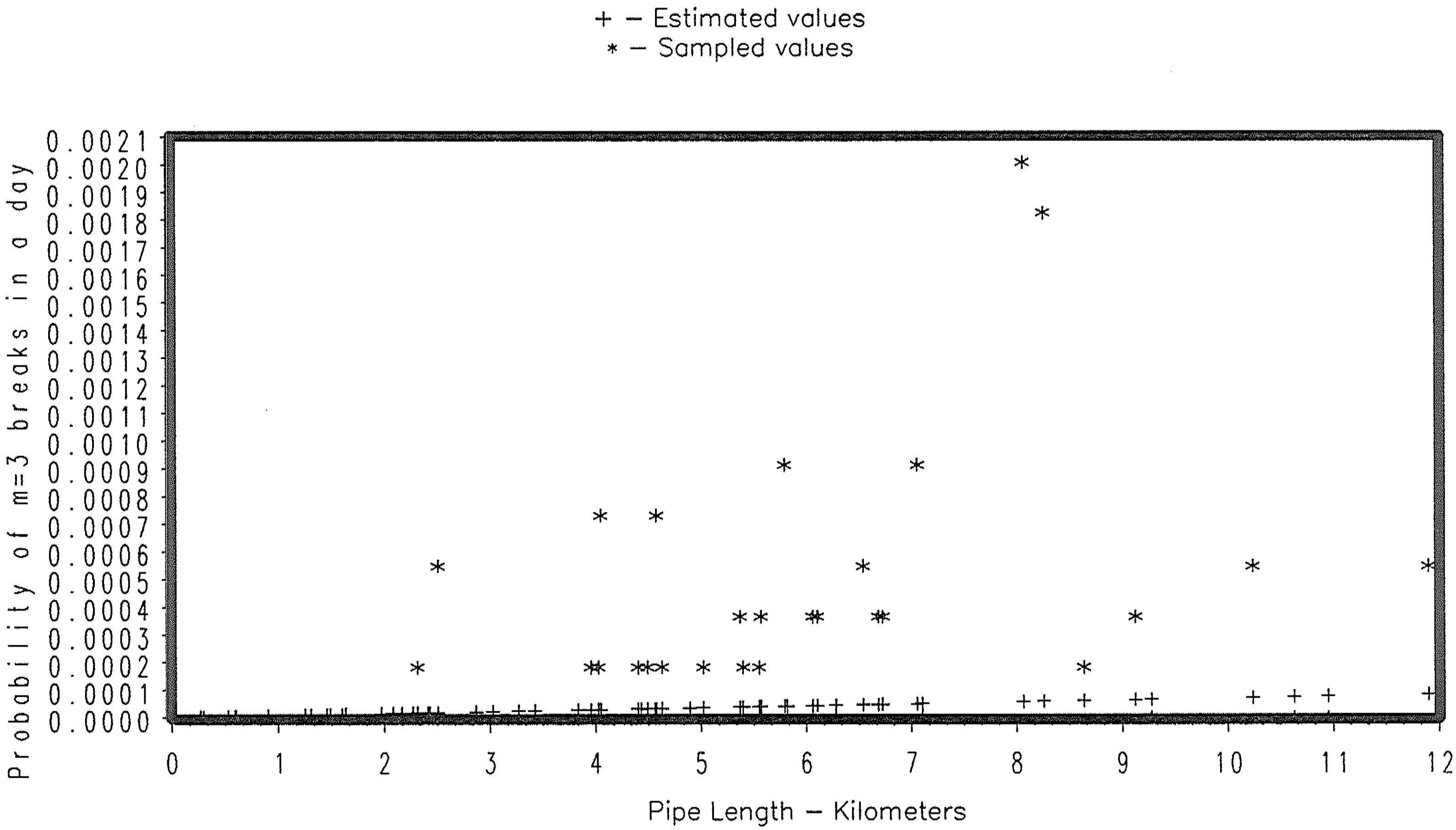


FIGURE B.93 Predicted versus sampled for m=3 breaks: cast iron pipe, 300mm 31+ years

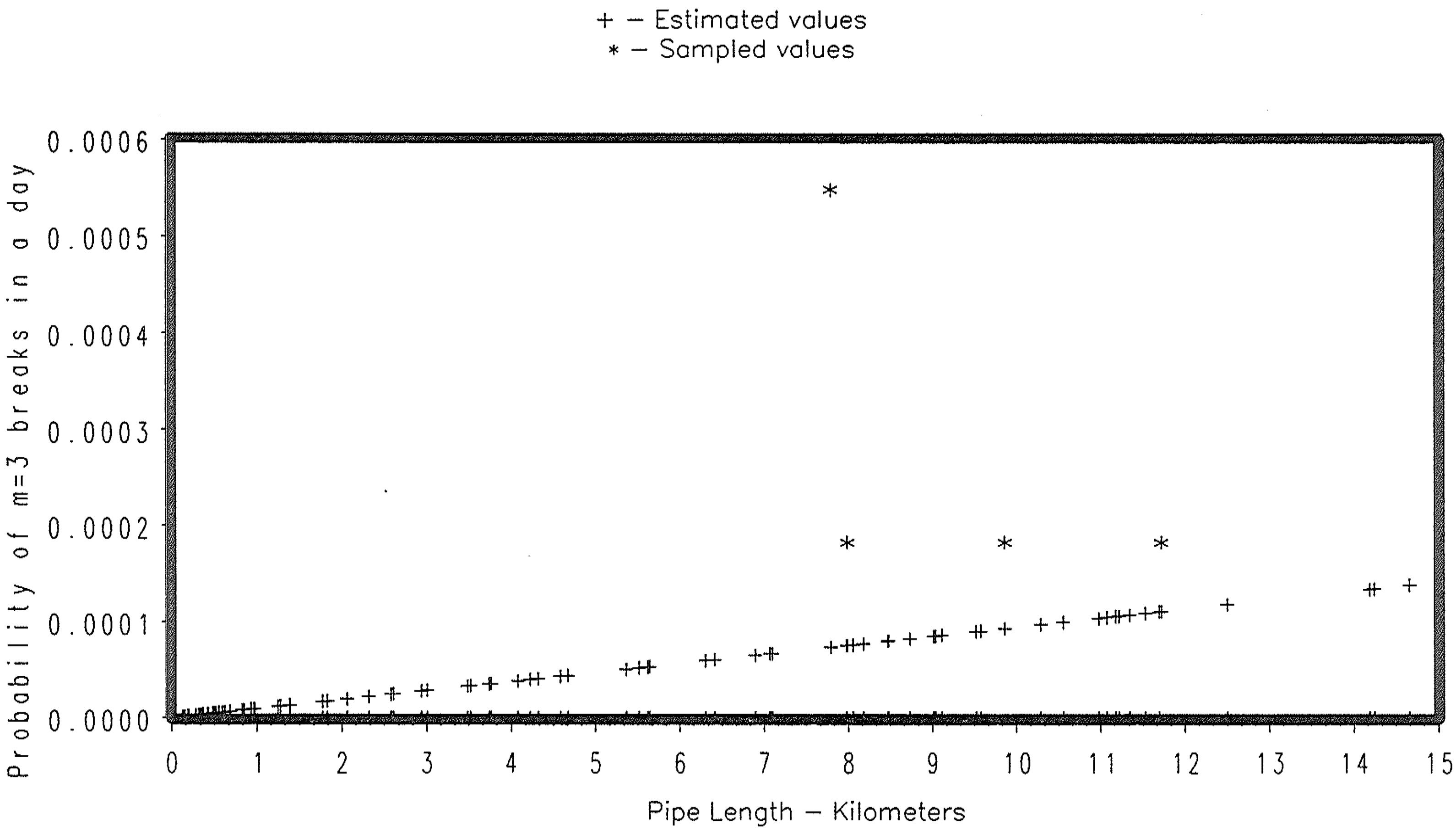


FIGURE B.94 Predicted versus sampled for m=3 breaks: asbestos cement, 150mm 0–30 years

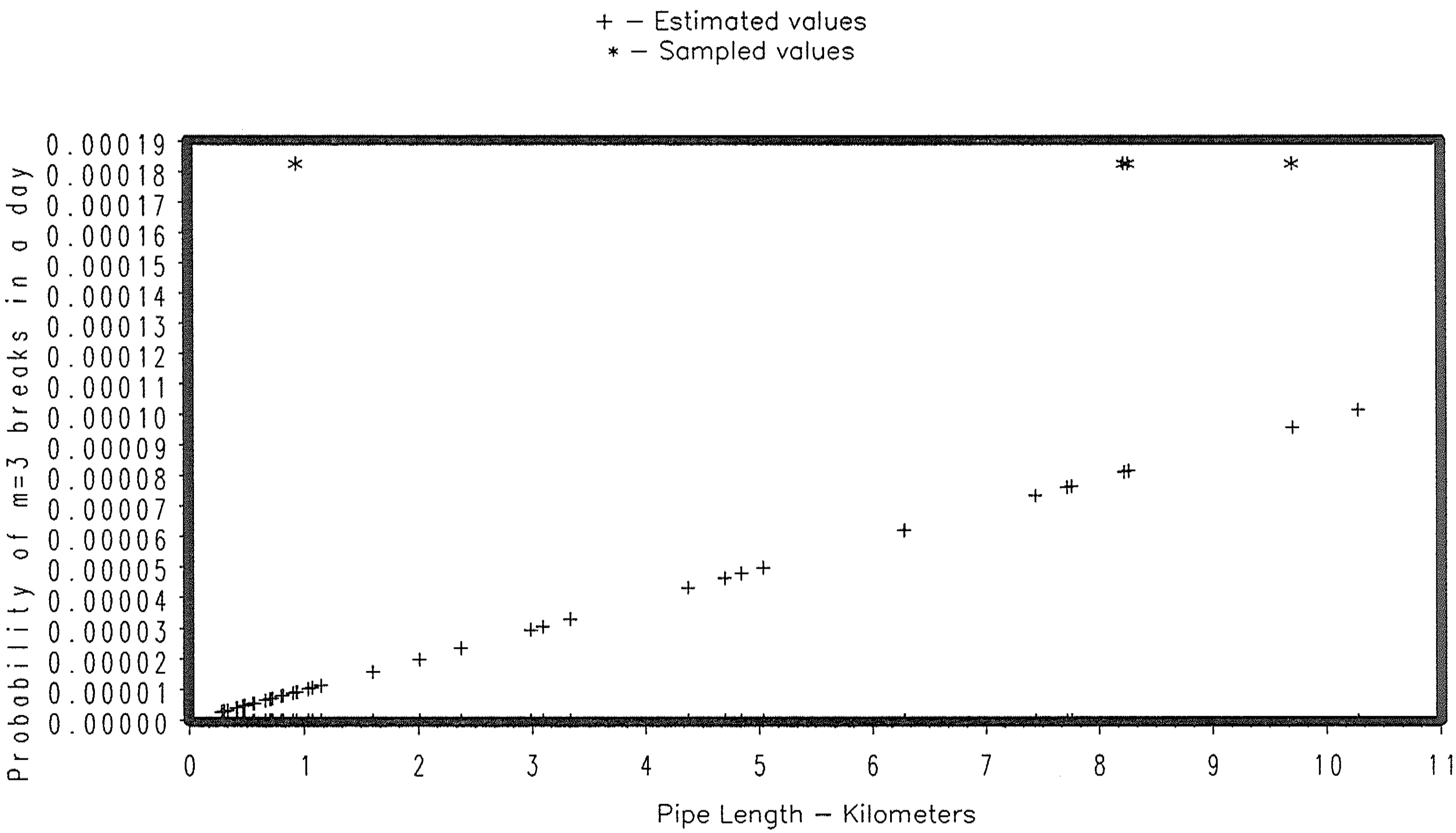


FIGURE B.95 Predicted versus sampled for m=3 breaks: asbestos cement, 150mm 31+ year

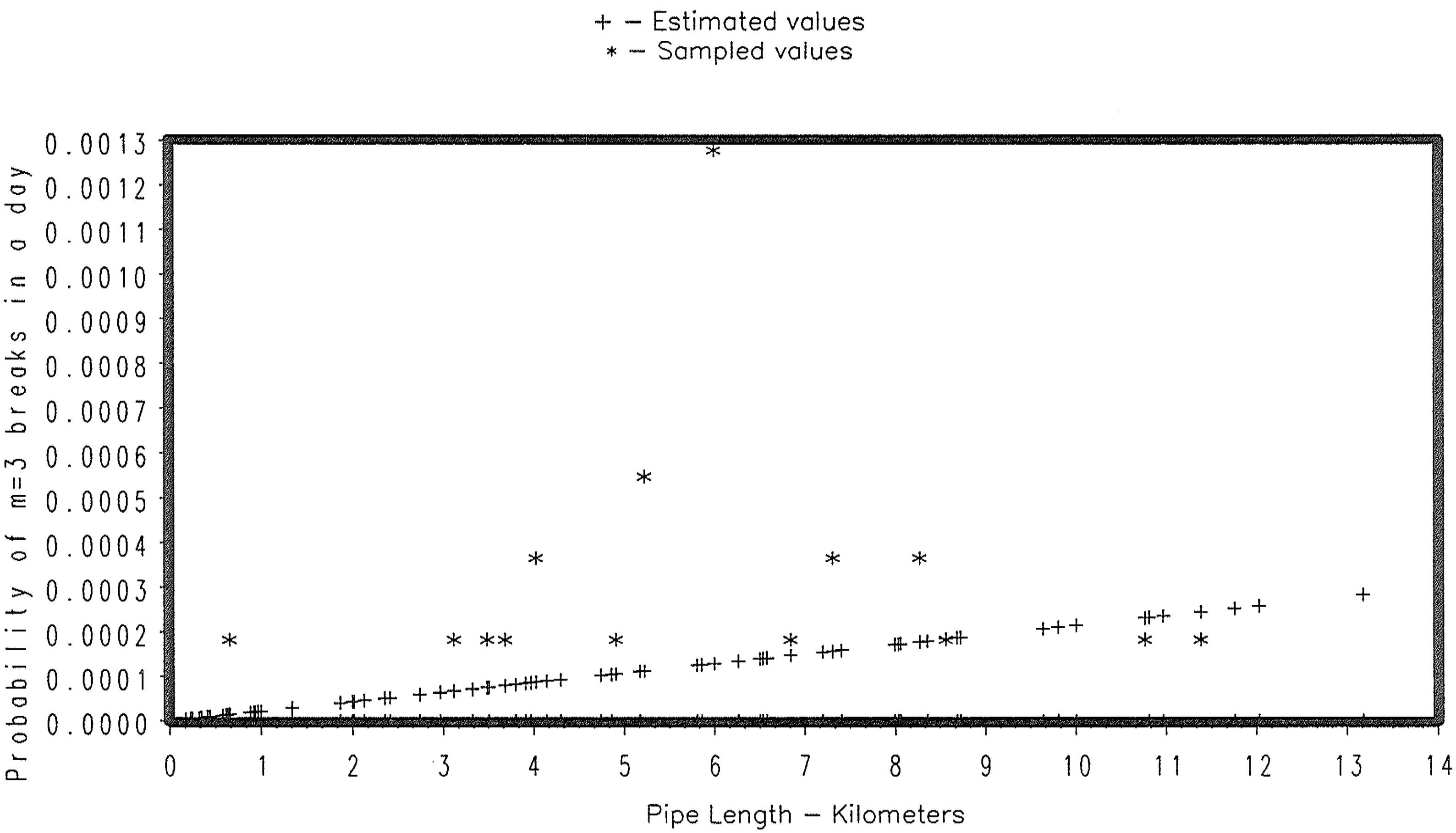


FIGURE B.96 Predicted versus sampled for m=3 breaks: asbestos cement, 200mm 0–30 years

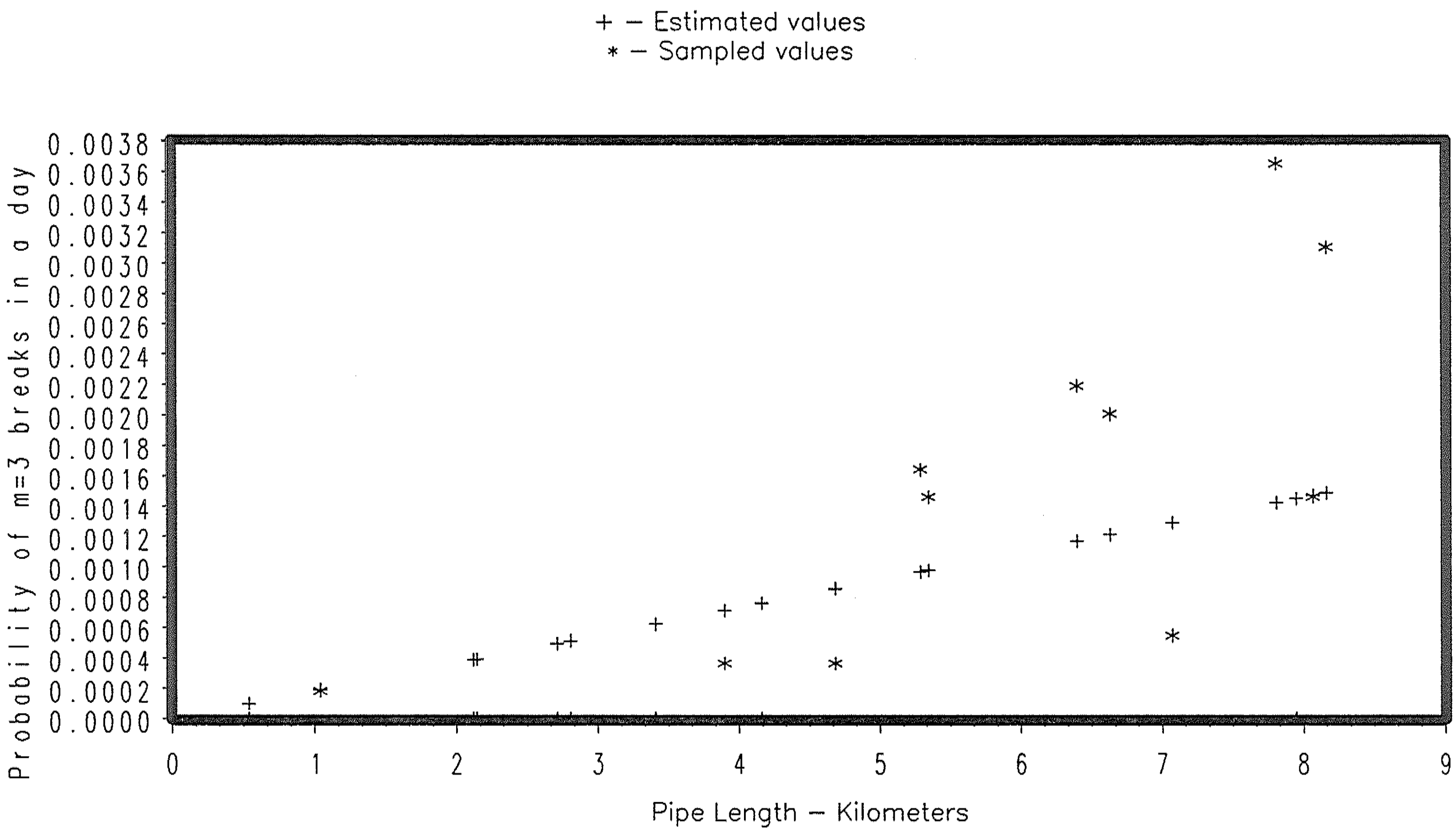


FIGURE B.97 Predicted versus sampled for m=3 breaks: asbestos cement, 200mm 31+ years

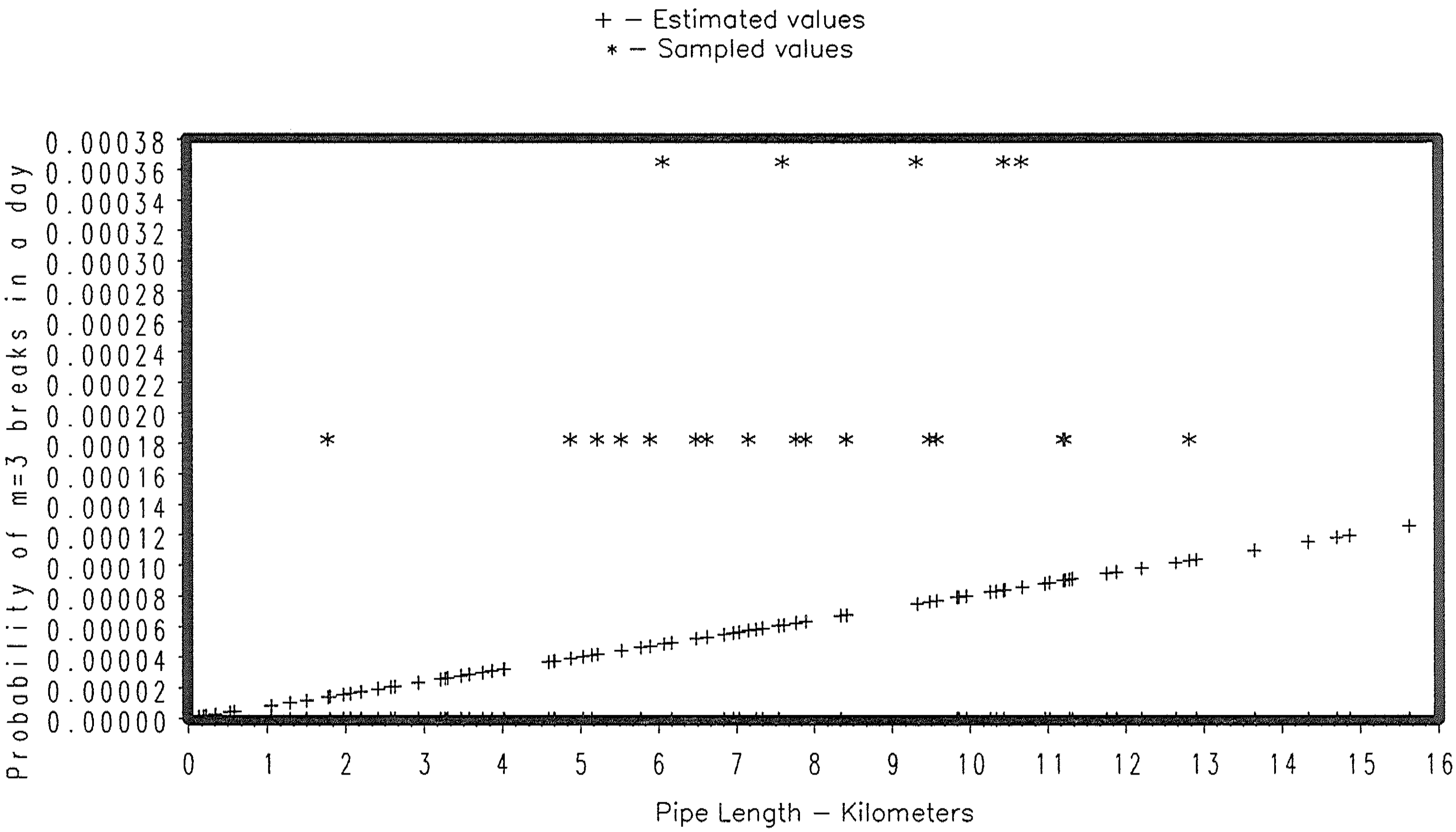


FIGURE B.98 Predicted versus sampled for m=3 breaks: asbestos cement, 250mm 0+ years

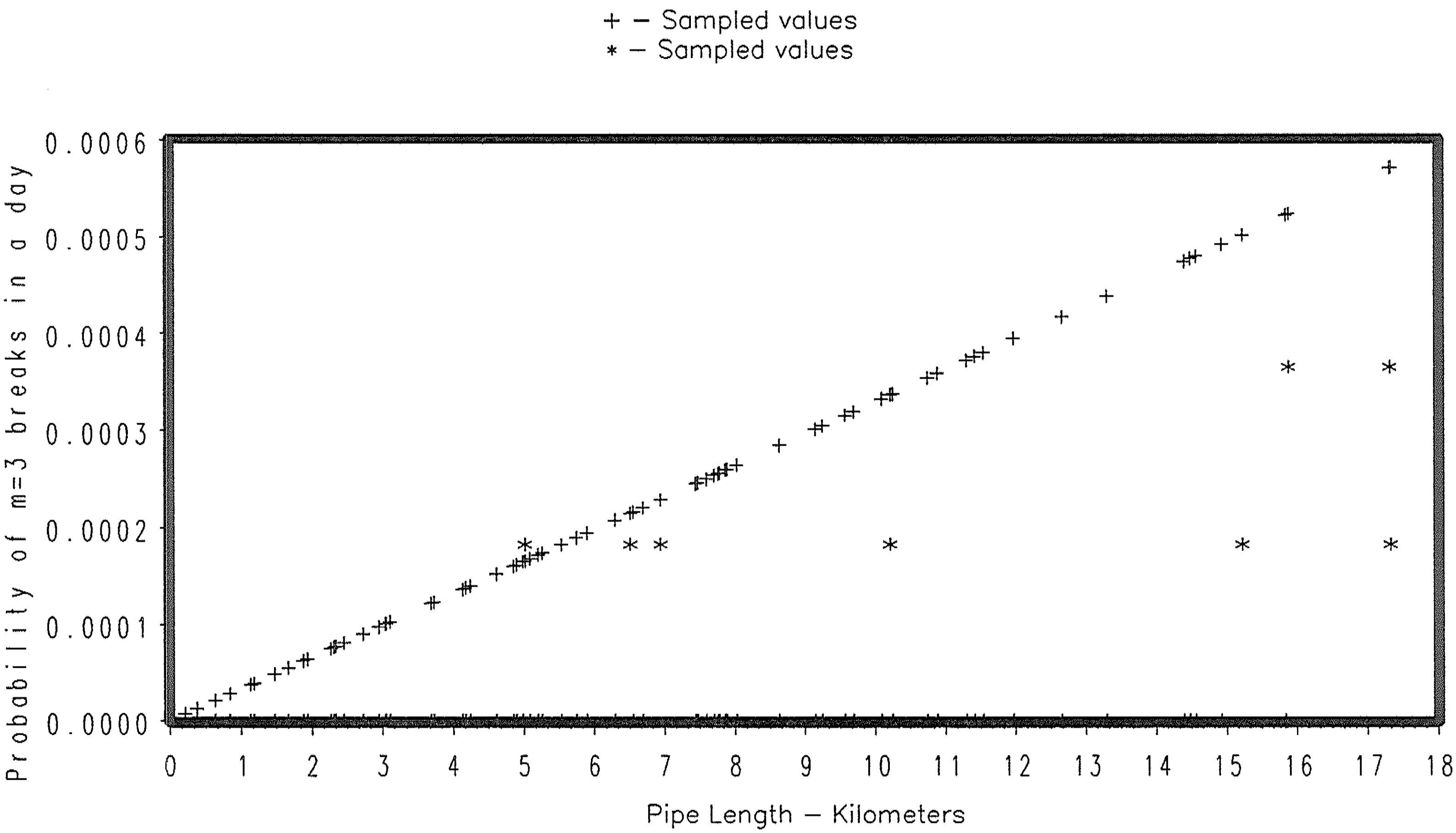


FIGURE B.99 Predicted versus sampled for m=3 breaks: asbestos cement, 300mm 0+ years

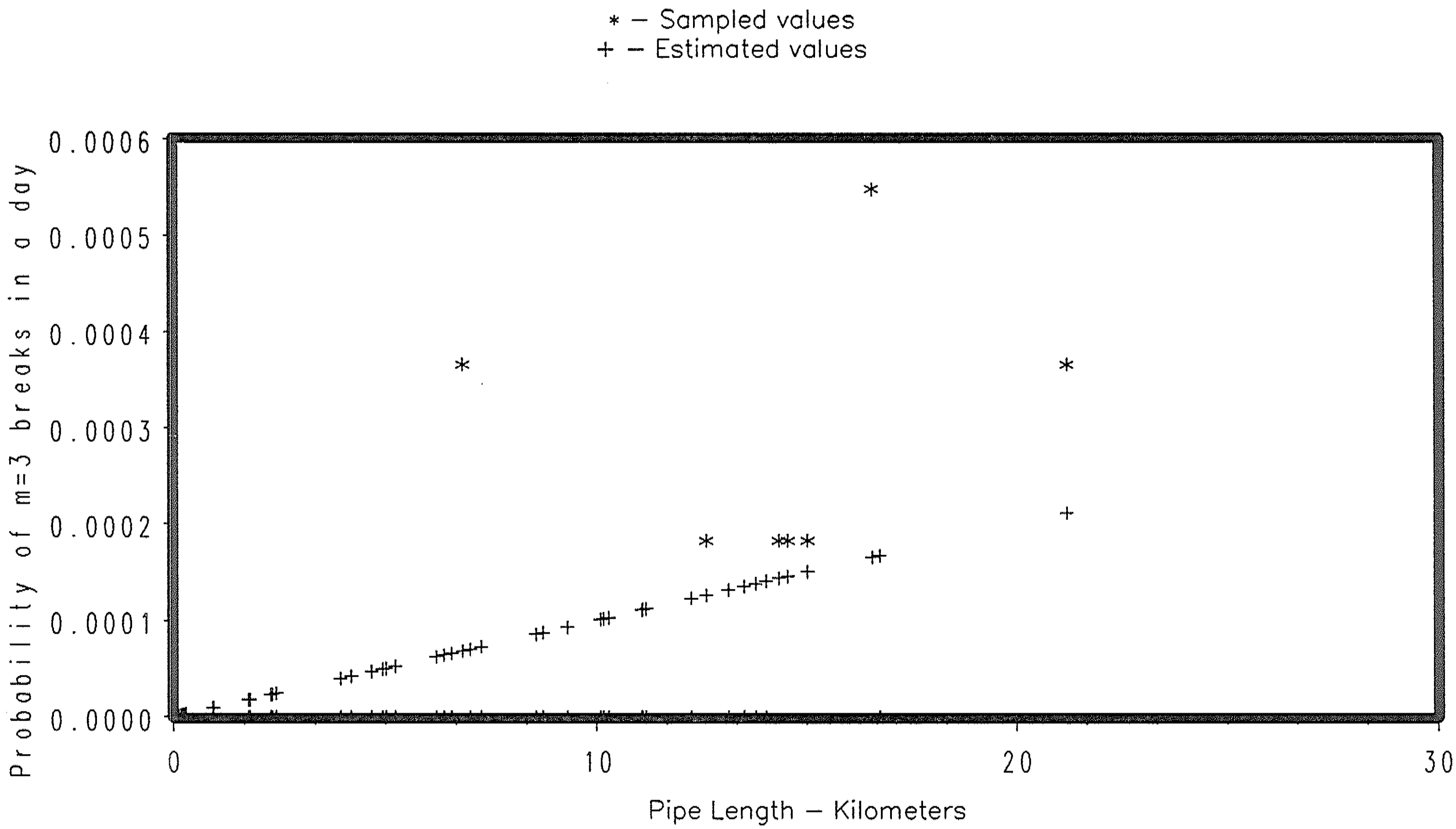


FIGURE B.100 Predicted versus sampled for m=3 breaks: ductile steel, 150mm 0+ years

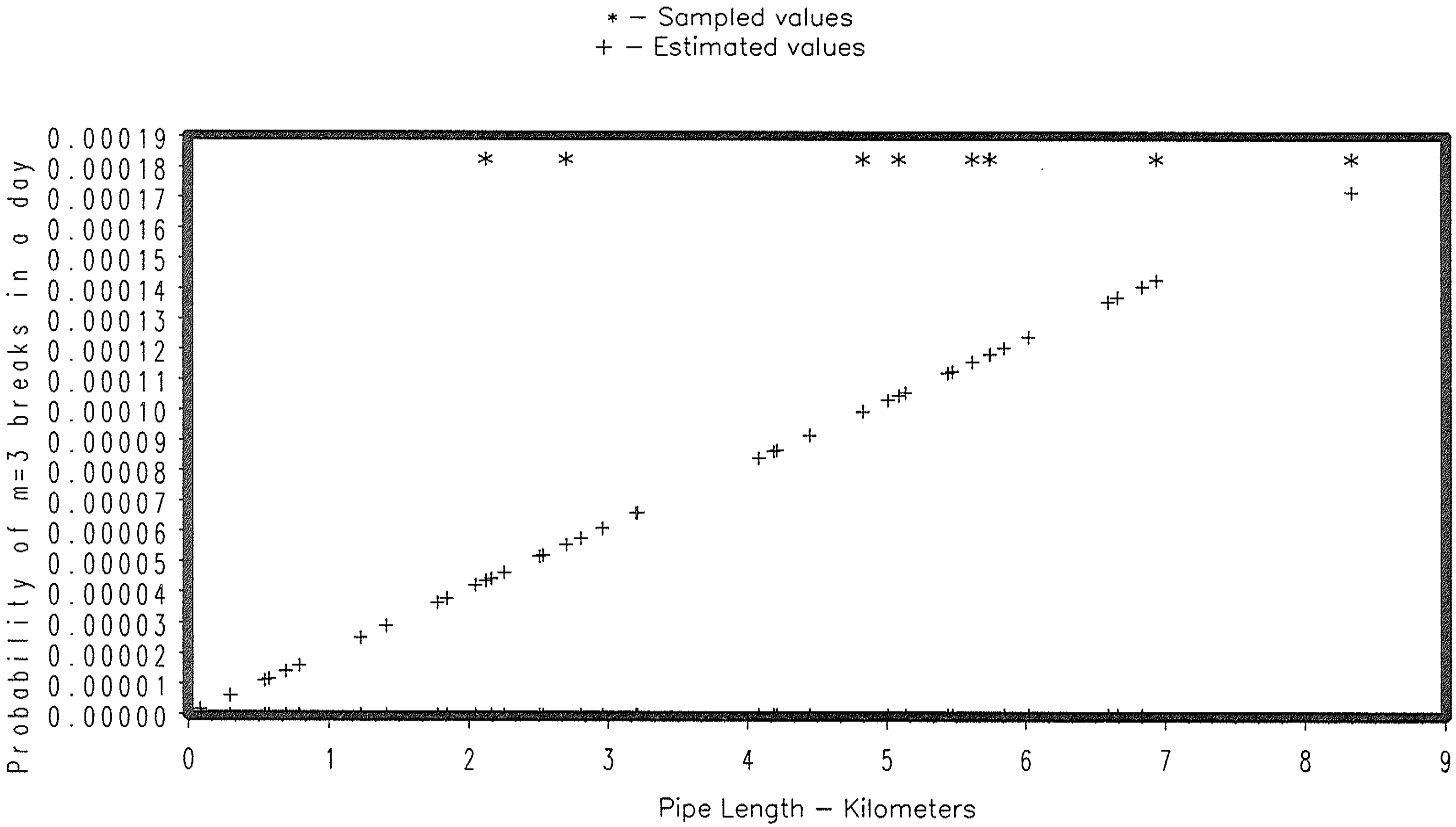


FIGURE B.101 Predicted versus sampled for m=3 breaks: ductile steel, 200mm 0+ years

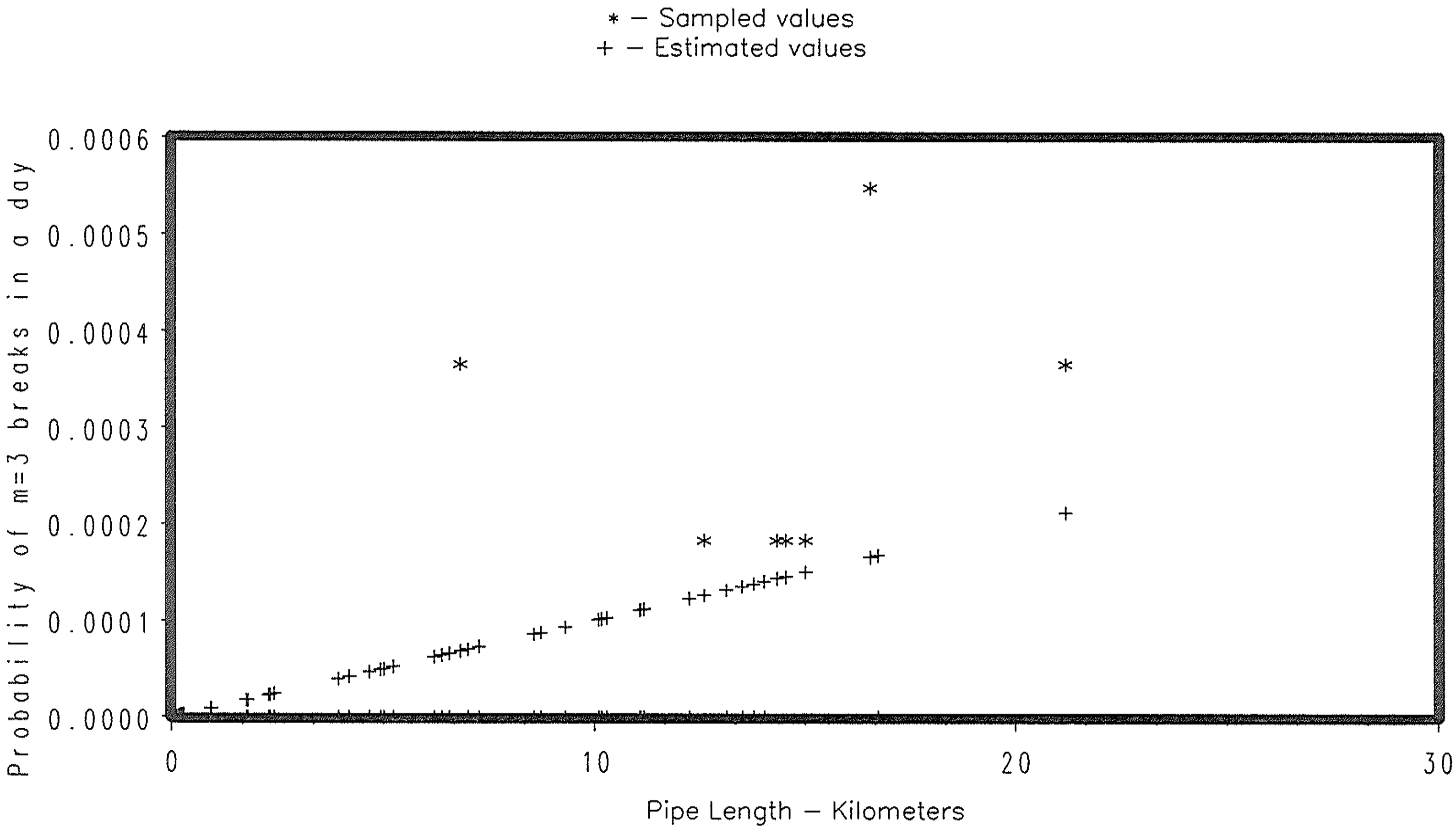


FIGURE B.102 Predicted versus sampled for m=3 breaks: ductile steel, 250mm 0+ years

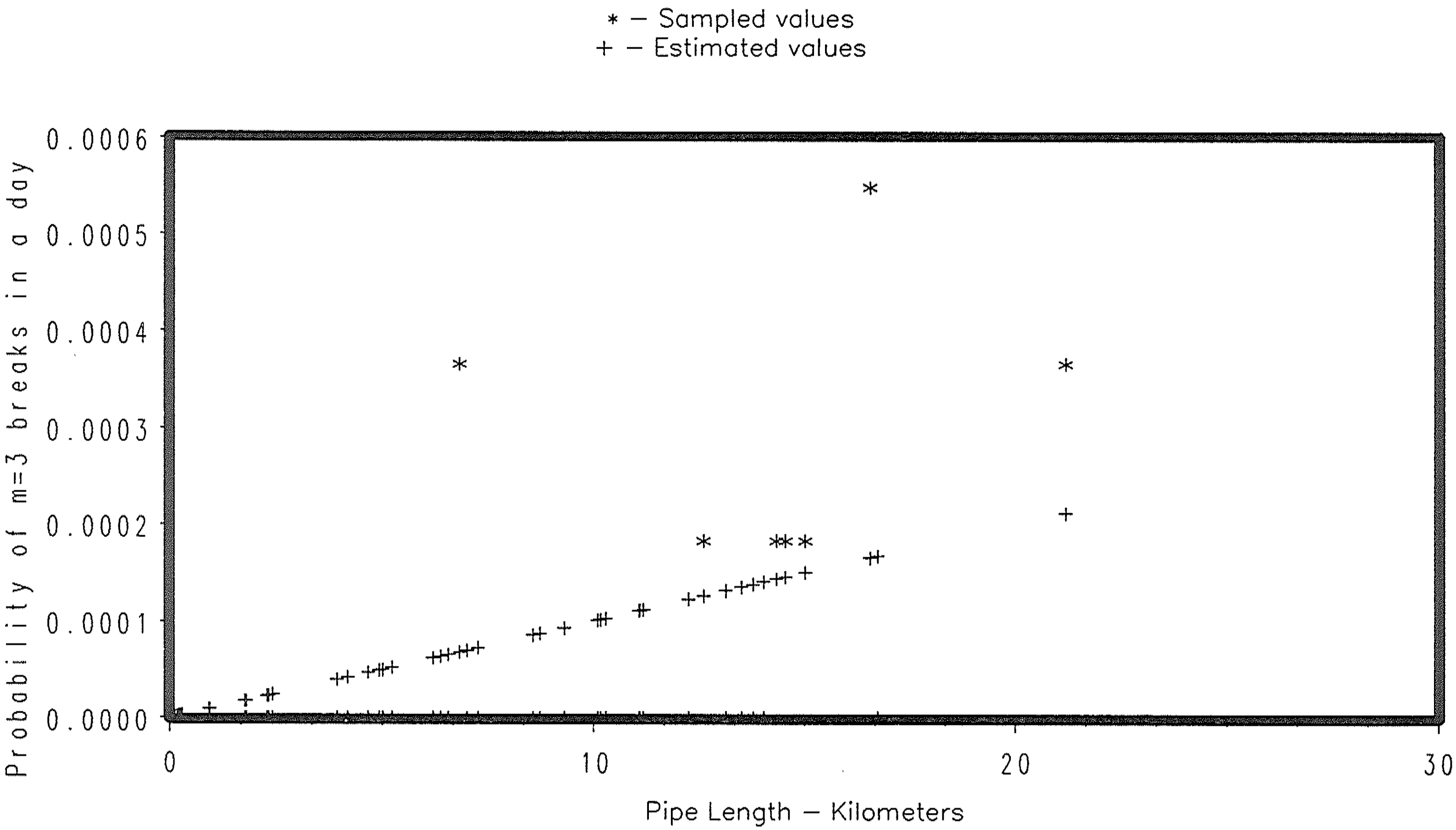


FIGURE B.103 Predicted versus sampled for m=3 breaks: PVC, 150mm 0+ years