



Statistical Modelling and Analysis of Gene Expression Data

Muhamed Wael Farouq
Teaching Assistant
Statistics, Mathematics & Insurance Department
Faculty of Commerce
Ain Shams University

A dissertation submitted in partial fulfilment of the
requirements for the degree of Doctor of Philosophy
(Applied Statistics)

Under Supervision of:

Professor Mostafa Galal Mostafa

Statistics, Mathematics
& Insurance Department
Faculty of Commerce
Ain Shams University

Professor Medhat Abdel-Aal

Statistics, Mathematics
& Insurance Department
Faculty of Commerce
Ain Shams University

Professor Abdel-Badeeh Salem

Faculty of Computer & Information Sciences
Ain Shams University

2018



Approval Sheet

Title of Thesis : Statistical Modeling and Analysis of Gene Expression Data

Academic Degree : Ph.D. in Applied Statistics

Name of Student : Muhamed Wael Farouq Abdel-Azeim

This thesis submitted in partial fulfillment of the requirements for the Degree of Philosophy in Applied Statistics has been approved by:

Examination committee

1- Professor Dr. Mustafa Galal Mustafa

Professor – Mathematics & Statistics Department

Faculty of Commerce

Ain Shams university

2- Professor Dr. Medhat Mohamed Abdel-Aal

Professor – Mathematics & Statistics Department

Faculty of Commerce

Ain Shams University

3- Professor Dr. Abdel-Badeeh Mohamed Salem

Professor - Faculty of Computer and Information Sciences

Ain Shams University

4- Professor Dr. Abdulla Ahmed Abd El-Ghaly

Professor - Faculty of Economics and Political Sciences

Cairo University

5- Dr. Mamdouh Abdel-Aleem

Associate Professor – Mathematics & Statistics Department

Faculty of Commerce

Ain Shams University

Date of Dissertation Defense: 29/12/2018 Approval date / /

Acknowledgements

I would like to express my deep appreciation to Prof. Mustafa Galal Mustafa for his support, encouragement, and patience during my years of study.

I would also like to express my gratitude to my supervisor Prof. Medhat Mohamed Abdelaal for allowing me to work on this project. I have learned a lot from him, his constructive appraisal has kept me motivated and interested in this work the whole time.

My deepest appreciation goes to my supervisor, Prof. Amir Hussain at University of Stirling, UK for his supervision, generous support and advice in all the aspects of this work. I sincerely thank him for providing me the opportunity for the joint supervision of my PhD at University of Stirling, UK. It would have never been achievable for me to take this work to completion without his constant feedback and answering my queries promptly.

I sincerely acknowledge and express my gratitude to Dr. Jozsef Farakas at University of Stirling, UK for his co-supervision. For him goes the credit for enlightening me the first glance of research in the mathematical aspects of this work. As well as for his insightful discussions and weekly follow-up meetings.

I will forever be thankful to Dr. Wadii Boulila. Wadii had confidence in my research and has been a fount of confidence and inspiration to me. His help was invaluable at many stages

in the course of this research project beside his careful editing contributed enormously to the production of this thesis.

My gratitude also goes to Prof. Abdel-Badeeh Mohamed Salem for the guidance in all aspects of the research, patience and ongoing support through the time of the PhD. I also have to thank the members of my PhD committee, Prof. Abdulla Ahmed Abd El-Ghaly and Dr. Mamdouh Abdel-aleem for their helpful career advice, valuable time and patience.

I gratefully acknowledge the funding received towards my PhD from the Egyptian Cultural Affairs and Mission Sector to pursue my joint supervision PhD at University of Stirling, UK.

I would like also to thank all my colleagues and members of staff at Department of Computing Science and Mathematics at University of Stirling, UK. They willingly lent me their time, resources and expertise. They expressed interest in the project and provided formal and informal constructive feedback on the research. Besides, their friendly environment and hospitality has made the beautiful city of Stirling more charming.

I also wish to take this opportunity to thank my wife, Nagwan and Maya for their love, support and entertainment. Last but not least, I am indebted to my parents, brothers and sisters for all their support and comfort. My parents have sacrificed their lives for my sisters, brothers and myself and provided unconditional love and care. I love them so much. I surely wouldn't have come this far without them.

To the soul of my father....

I love you.

Muhamed Wael Farouq

Table of Contents

Chapter One: Introduction		
1.1	Purpose of the Research	2
1.2	Significance of the Research	3
1.3	Research Hypotheses	4
1.4	Limitation of the Research	5
1.5	Dataset	5
1.6	Plan of the Research	6
Chapter Two: Biological Background		
2.1	Introduction	7
2.2	Gene expression via RNA	9
2.3	Epigenetics	10
2.4	Small Non-coding RNA (ncRNA)	15
2.5	Cancer Genomics	19
2.6	Representation of Gene Expression Data	21
Chapter Three: Clustering and Dempster-Shafer Theory		
3.1	Introduction	23
3.2	Clustering	23
3.3	Clustering Tendency	26
3.4	Criteria of Clustering	29
3.5	Optimisation Algorithms	30
3.6	Proximity Measures	33
3.7	Clustering Objective Function	40
3.8	Validation and Cardinality of Clustering	43
3.9	Ensemble of Clustering Validity Indices	61
3.10	Post processing Analysis of Clustering	62
3.11	Literature Review of Clustering Algorithms for Gene Expression	63
3.12	Evidence Theory and Dempster-Shafer	66
3.13	Literature Review of Evidence Theory and Gene Expression	68
3.14	Proposed Framework	70

Table of Contents - Continued

Chapter Four: Mathematical Modelling of Population Dynamics		
4.1	Introduction	75
4.2	First Order Ordinary Differential Equation The Euler Method	81
4.3	First Order Ordinary Differential Equation Separable Equations	82
4.4	Second Order Ordinary Differential Equation The Euler Method	83
4.5	Analytical Solution for the Second Order Linear Homogenous ODE with Constant Coefficients	85
4.6	From Ordinary Differential Equations to Partial Differential Equations	89
4.7	Analytical Solution of the First Order Linear PDE with Constant Coefficients	94
4.8	Second Order Linear Homogenous PDE with Constant Coefficient	96
4.9	Diffusion Equation	104
Chapter Five: Computational Results and Qualitative Analysis		
5.1	Two Species Diffusion Equation	119
5.2	Two species Model in Dirichlet Initial Boundary Value Problem	125
5.3	Two species Model in Neumann Initial Boundary Value Problem	128
5.4	Interpretation	128
5.5	Experimental Results	175
Chapter Six: Conclusion and Future Work		
6.1	Conclusion	185
6.2	Future Work	186

List of Figures

Figure 2.1: DNA Molecule	8
Figure 2.2: mRNA Transcription	10
Figure 2.3: Gene Dynamic Scheme	13
Figure 2.4: DNA Methylation	15
Figure 2.5: Pre-transcription, Transcription and Post-Transcription Scheme	18
Figure 3.1: Evidence Structure	68
Figure 3.2: Proposed Framework	72
Figure 3.3: Fusion Clustering Scheme	74
Figure 5.1: 2D Initial Condition Consistency Correction	127
Figure 5.2: 3D Initial Condition Consistency Correction	127
Figure 5.3a: Two Species Model 1.1 mRNA Population at Transition States	132
Figure 5.3b: Two Species Model 1.1 Total mRNA Population at Transition States	132
Figure 5.4a: Two Species Model 1.2 mRNA Population at Transition States	132
Figure 5.4b: Two Species Model 1.2 Total mRNA Population at Transition States	132
Figure 5.5a: Two Species Model 1.3 mRNA Population at Transition States	133
Figure 5.5b: Two Species Model 1.3 Total mRNA Population at Transition States	133
Figure 5.6a: Two Species Model 1.4 mRNA Population at Transition States	133
Figure 5.6b: Two Species Model 1.4 Total mRNA Population at Transition States	133
Figure 5.7a: Two Species Model 1.5 mRNA Population at Transition States	134
Figure 5.7b: Two Species Model 1.5 Total mRNA Population at Transition States	134
	134

Figure 5.8a: Two Species Model 1.6 mRNA Population at Transition States		
Figure 5.8b: Two Species Model 1.6 Total mRNA Population at Transition States	134
Figure 5.9a: Two Species Model 1.7 mRNA Population at Transition States	135
Figure 5.9b: Two Species Model 1.7 Total mRNA Population at Transition States	135
Figure 5.10a: Two Species Model 1.8 mRNA Population at Transition States	135
Figure 5.10b: Two Species Model 1.8 Total mRNA Population at Transition States	135
Figure 5.11a: Two Species Model 1.9 mRNA Population at Transition States	136
Figure 5.11b: Two Species Model 1.9 Total mRNA Population at Transition States	136
Figure 5.12a: Two Species Model 1.10 mRNA Population at Transition States	136
Figure 5.12b: Two Species Model 1.10 Total mRNA Population at Transition States	136
Figure 5.13a: Two Species Model 1.11 mRNA Population at Transition States	137
Figure 5.13b: Two Species Model 1.11 Total mRNA Population at Transition States	137
Figure 5.14a: Two Species Model 1.12 mRNA Population at Transition States	137
Figure 5.14b: Two Species Model 1.12 Total mRNA Population at Transition States	137
Figure 5.15a: Two Species Model 1.13 mRNA Population at Transition States	138
Figure 5.15b: Two Species Model 1.13 Total mRNA Population at Transition States	138
Figure 5.16a: Two Species Model 1.14 mRNA Population at Transition States	138
Figure 5.16b: Two Species Model 1.14 Total mRNA Population at Transition States	138
Figure 5.17a: Two Species Model 1.15 mRNA Population at Transition States	139
Figure 5.17b: Two Species Model 1.15 Total mRNA Population at Transition States	139

Figure 5.18a: Two Species Model 1.16 mRNA Population at Transition States	139
Figure 5.18b: Two Species Model 1.16 Total mRNA Population at Transition States	139
Figure 5.19a: Two Species Model 1.17 mRNA Population at Transition States	140
Figure 5.19b: Two Species Model 1.17 Total mRNA Population at Transition States	140
Figure 5.20a: Two Species Model 1.18 mRNA Population at Transition States	140
Figure 5.20b: Two Species Model 1.18 Total mRNA Population at Transition States	140
Figure 5.21a: Two Species Model 1.19 mRNA Population at Transition States	141
Figure 5.21b: Two Species Model 1.19 Total mRNA Population at Transition States	141
Figure 5.22a: Two Species Model 2.1 mRNA Population at Transition States	143
Figure 5.22b: Two Species Model 2.1 Total mRNA Population at Transition States	143
Figure 5.23a: Two Species Model 2.2 mRNA Population at Transition States	143
Figure 5.23b: Two Species Model 2.2 Total mRNA Population at Transition States	143
Figure 5.24a: Two Species Model 2.3 mRNA Population at Transition States	144
Figure 5.24b: Two Species Model 2.3 Total mRNA Population at Transition States	144
Figure 5.25a: Two Species Model 2.4 mRNA Population at Transition States	144
Figure 5.25b: Two Species Model 2.4 Total mRNA Population at Transition States	144
Figure 5.26a: Two Species Model 2.5 mRNA Population at Transition States	145
Figure 5.26b: Two Species Model 2.5 Total mRNA Population at Transition States	145
Figure 5.27a: Two Species Model 2.6 mRNA Population at Transition States	145
Figure 5.27b: Two Species Model 2.6 Total mRNA Population at Transition States	145

Figure 5.28a: Two Species Model 2.7 mRNA Population at Transition States	146
Figure 5.28b: Two Species Model 2.7 Total mRNA Population at Transition States	146
Figure 5.29a: Two Species Model 2.8 mRNA Population at Transition States	146
Figure 5.29b: Two Species Model 2.8 Total mRNA Population at Transition States	146
Figure 5.30a: Two Species Model 2.9 mRNA Population at Transition States	147
Figure 5.30b: Two Species Model 2.9 Total mRNA Population at Transition States	147
Figure 5.31a: Two Species Model 2.10 mRNA Population at Transition States	147
Figure 5.31b: Two Species Model 2.10 Total mRNA Population at Transition States	147
Figure 5.32a: Two Species Model 2.11 mRNA Population at Transition States	148
Figure 5.32b: Two Species Model 2.11 Total mRNA Population at Transition States	148
Figure 5.33a: Two Species Model 2.12 mRNA Population at Transition States	148
Figure 5.33b: Two Species Model 2.12 Total mRNA Population at Transition States	148
Figure 5.34a: Two Species Model 2.13 mRNA Population at Transition States	149
Figure 5.34b: Two Species Model 2.13 Total mRNA Population at Transition States	149
Figure 5.35a: Two Species Model 2.14 mRNA Population at Transition States	149
Figure 5.35b: Two Species Model 2.14 Total mRNA Population at Transition States	149
Figure 5.36a: Two Species Model 2.15 mRNA Population at Transition States	150
Figure 5.36b: Two Species Model 2.15 Total mRNA Population at Transition States	150
Figure 5.37a: Two Species Model 2.16 mRNA Population at Transition States	150
Figure 5.37b: Two Species Model 2.16 Total mRNA Population at Transition States	150

Figure 5.38a: Two Species Model 2.17 mRNA Population at Transition States	151
Figure 5.38b: Two Species Model 2.17 Total mRNA Population at Transition States	151
Figure 5.39a: Two Species Model 2.18 mRNA Population at Transition States	151
Figure 5.39b: Two Species Model 2.18 Total mRNA Population at Transition States	151
Figure 5.40a: Two Species Model 2.19 mRNA Population at Transition States	152
Figure 5.40b: Two Species Model 2.19 Total mRNA Population at Transition States	152
Figure 5.41a: Two Species Model 3.1 mRNA Population at Transition States	154
Figure 5.41b: Two Species Model 3.1 Total mRNA Population at Transition States	154
Figure 5.42a: Two Species Model 3.2 mRNA Population at Transition States	154
Figure 5.42b: Two Species Model 3.2 Total mRNA Population at Transition States	154
Figure 5.43a: Two Species Model 3.3 mRNA Population at Transition States	155
Figure 5.43b: Two Species Model 3.3 Total mRNA Population at Transition States	155
Figure 5.44a: Two Species Model 3.4 mRNA Population at Transition States	155
Figure 5.44b: Two Species Model 3.4 Total mRNA Population at Transition States	155
Figure 5.45a: Two Species Model 3.5 mRNA Population at Transition States	156
Figure 5.45b: Two Species Model 3.5 Total mRNA Population at Transition States	156
Figure 5.46a: Two Species Model 3.6 mRNA Population at Transition States	156
Figure 5.46b: Two Species Model 3.6 Total mRNA Population at Transition States	156
Figure 5.47a: Two Species Model 3.7 mRNA Population at Transition States	157
Figure 5.47b: Two Species Model 3.7 Total mRNA Population at Transition States	157

Figure 5.48a: Two Species Model 3.8 mRNA Population at Transition States	157
Figure 5.48b: Two Species Model 3.8 Total mRNA Population at Transition States	157
Figure 5.49a: Two Species Model 3.9 mRNA Population at Transition States	158
Figure 5.49b: Two Species Model 3.9 Total mRNA Population at Transition States	158
Figure 5.50a: Two Species Model 3.10 mRNA Population at Transition States	158
Figure 5.50b: Two Species Model 3.10 Total mRNA Population at Transition States	158
Figure 5.51a: Two Species Model 3.11 mRNA Population at Transition States	159
Figure 5.51b: Two Species Model 3.11 Total mRNA Population at Transition States	159
Figure 5.52a: Two Species Model 3.12 mRNA Population at Transition States	159
Figure 5.52b: Two Species Model 3.12 Total mRNA Population at Transition States	159
Figure 5.53a: Two Species Model 3.13 mRNA Population at Transition States	160
Figure 5.53b: Two Species Model 3.13 Total mRNA Population at Transition States	160
Figure 5.54a: Two Species Model 3.14 mRNA Population at Transition States	160
Figure 5.54b: Two Species Model 3.14 Total mRNA Population at Transition States	160
Figure 5.55a: Two Species Model 3.15 mRNA Population at Transition States	161
Figure 5.55b: Two Species Model 3.15 Total mRNA Population at Transition States	161
Figure 5.56a: Two Species Model 3.16 mRNA Population at Transition States	161
Figure 5.56b: Two Species Model 3.16 Total mRNA Population at Transition States	161
Figure 5.57a: Two Species Model 3.17 mRNA Population at Transition States	162
Figure 5.57b: Two Species Model 3.17 Total mRNA Population at Transition States	162

Figure 5.58a: Two Species Model 3.18 mRNA Population at Transition States	162
Figure 5.58b: Two Species Model 3.18 Total mRNA Population at Transition States	162
Figure 5.59a: Two Species Model 3.19 mRNA Population at Transition States	163
Figure 5.59b: Two Species Model 3.19 Total mRNA Population at Transition States	163
Figure 5.60a: Two Species Model 3.20 mRNA Population at Transition States	163
Figure 5.60b: Two Species Model 3.20 Total mRNA Population at Transition States	163
Figure 5.61a: Two Species Model 3.21 mRNA Population at Transition States	164
Figure 5.61b: Two Species Model 3.21 Total mRNA Population at Transition States	164
Figure 5.62a: Two Species Model 3.22 mRNA Population at Transition States	164
Figure 5.62b: Two Species Model 3.22 Total mRNA Population at Transition States	164
Figure 5.63a: Two Species Model 4.1 mRNA Population at Transition States	166
Figure 5.63b: Two Species Model 4.1 Total mRNA Population at Transition States	166
Figure 5.64a: Two Species Model 4.2 mRNA Population at Transition States	166
Figure 5.64b: Two Species Model 4.2 Total mRNA Population at Transition States	166
Figure 5.65a: Two Species Model 4.3 mRNA Population at Transition States	167
Figure 5.65b: Two Species Model 4.3 Total mRNA Population at Transition States	167
Figure 5.66a: Two Species Model 4.4 mRNA Population at Transition States	167
Figure 5.66b: Two Species Model 4.4 Total mRNA Population at Transition States	167
Figure 5.67a: Two Species Model 4.5 mRNA Population at Transition States	168
Figure 5.67b: Two Species Model 4.5 Total mRNA Population at Transition States	168

Figure 5.68a: Two Species Model 4.6 mRNA Population at Transition States	168
Figure 5.68b: Two Species Model 4.6 Total mRNA Population at Transition States	168
Figure 5.69a: Two Species Model 4.7 mRNA Population at Transition States	169
Figure 5.69b: Two Species Model 4.7 Total mRNA Population at Transition States	169
Figure 5.70a: Two Species Model 4.8 mRNA Population at Transition States	169
Figure 5.70b: Two Species Model 4.8 Total mRNA Population at Transition States	169
Figure 5.71a: Two Species Model 4.9 mRNA Population at Transition States	170
Figure 5.71b: Two Species Model 4.9 Total mRNA Population at Transition States	170
Figure 5.72a: Two Species Model 4.10 mRNA Population at Transition States	170
Figure 5.72b: Two Species Model 4.10 Total mRNA Population at Transition States	170
Figure 5.73a: Two Species Model 4.11 mRNA Population at Transition States	171
Figure 5.73b: Two Species Model 4.11 Total mRNA Population at Transition States	171
Figure 5.74a: Two Species Model 4.12 mRNA Population at Transition States	171
Figure 5.74b: Two Species Model 4.12 Total mRNA Population at Transition States	171
Figure 5.75a: Two Species Model 4.13 mRNA Population at Transition States	172
Figure 5.75b: Two Species Model 4.13 Total mRNA Population at Transition States	172
Figure 5.76a: Two Species Model 4.14 mRNA Population at Transition States	172
Figure 5.76b: Two Species Model 4.14 Total mRNA Population at Transition States	172
Figure 5.77a: Two Species Model 4.15 mRNA Population at Transition States	173
Figure 5.77b: Two Species Model 4.15 Total mRNA Population at Transition States	173