DYNAMIC SPATIAL-TEMPORAL POINT PROCESS MODELS VIA CONDITIONING

A Thesis presented to
the Faculty of the Graduate School
at the University of Missouri

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
JUSTIN OBWOGE OKENYE
Drs. Athanasios C. Micheas and Christopher K. Wikle; Thesis Supervisors
JULY 2021
The undersigned, appointed by the Dean of the Graduate School, have examined the
dissertation entitled:

**DYNAMIC SPATIAL-TEMPORAL POINT PROCESS MODELS VIA CONDITIONING**

presented by **Justin Obwoge Okenye**, a candidate for the degree of Doctor of Philosophy
and hereby certify that, in their opinion, it is worthy of acceptance.

______________________________
Dr. Athanasios Micheas

______________________________
Dr. Christopher Wikle

______________________________
Dr. Subharup Guha

______________________________
Dr. Scott Holan

______________________________
Dr. Neil Fox
Dedication

With love to my loving children Lynn Kemunto and Ian Omambia.
Acknowledgements

First and foremost, my sincere gratitude goes to my advisors Dr. Athanasios Micheas and Dr. Chris Wikle. I thank them most sincerely for their patience, excellent guidance, unfailing willingness to help (both material and financial) but most importantly for the knowledge I have received from them as a graduate student here at the University of Missouri, Columbia. Without their tremendous support, enthusiasm and inspiration this dissertation would have been utterly impossible. They have been excellent mentors and will forever remain my inspiration with a solemn pledge on my part never to let them down.

I would also like to thank my committee members, Dr. Subharup Guha, Dr. Scott Holan and Dr. Neil Fox from whom I gained great knowledge. I will be forever grateful to all faculty members and staff of the Department Statistics and all my friends for all their support, helpful comments and advice throughout my stay here at Columbia. I most sincerely thank Judy Dooley and Kathleen Maurer for their helpful assistance on various occasions.

I wish to thank my parents Mr. Zedekiah Okenye Onyinkwa and Mrs. Jemimah Moraa Okenye and all my siblings; George Morara, James Moracha, Alfred Osoro and Dephine Nyahoke for all their love, sacrifice and great moral support throughout my life.

My special thanks go to my dear wife, Gladys Moraa and our two lovely children Lynn Kemunto (Magokoro) and Ian Omambia (Tata moke) for their emotional support and patience while I was away pursuing my Ph.D. They provided me with the strength and perseverance without which this work would have not been possible. I am eternally indebted to them for their love and prayers that constantly motivate me to accomplish my goals in my career.

It would be extremely unfair of me to fail to recognize the selflessly tireless efforts of my bosom friend Dr. Ogada Elisha (Jalolo, Ja Karachuonyo, my brother from another mother). This manuscript might not have been the way it is without the technical support, encouragement and insight from this friend of mine. Behind this document is a story of many difficult sleepless nights as we walked together in this journey to see this work afoot.
May God bless this son of Africa.

I will hardly forget my friend, Mr. Mohammed (Adam) Bouras, the lion from Atlas Mountains, and his brother. They were my wonderful hosts even as I tried to gain my footing in and around Colombia Missouri. May Allah shower them with blessings.

I would be remiss if I do not acknowledge the support I received from great people and Government of the United States through the State Department, Institute of International Education (IIE) and Fulbright scholarship board for giving the opportunity to study through the Fulbright Fellowship.
Contents

Acknowledgements ii

List of Figures viii

List of Tables xv

Abstract xxi

1 Modeling Point Patterns 1

1.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

1.2 Models for independent events . . . . . . . . . . . . . . . . . . . . . . . . . 4

1.2.1 Poisson point processes . . . . . . . . . . . . . . . . . . . . . . . . . . 5

1.2.2 Marked Poisson point processes . . . . . . . . . . . . . . . . . . . . . 6

1.2.3 Space-time Poisson point processes . . . . . . . . . . . . . . . . . . . . 8

1.2.4 Marked space-time Poisson point processes . . . . . . . . . . . . . . . 11

1.2.5 Why should we use conditioning? . . . . . . . . . . . . . . . . . . . . . 13

1.3 Models for interacting events . . . . . . . . . . . . . . . . . . . . . . . . . 13

1.3.1 Markov point process . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14

1.3.2 Marked Markov point process . . . . . . . . . . . . . . . . . . . . . . . 18

1.3.3 Space-time Markov point process . . . . . . . . . . . . . . . . . . . . . 19

1.4 Mixture models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20

1.4.1 Modeling the intensity function as a mixture . . . . . . . . . . . . . . . 22
2 Dynamic Spatio-Temporal Poisson Point Process Models via Conditioning on Time

2.1 Introduction ................................................................. 25

2.2 Modeling a spatial-temporal point process via conditioning .................................. 29

2.2.1 Marked space-time non-homogeneous Poisson point process models ........ 29

2.2.2 A multivariate space-time modeling approach ............................................. 38

2.3 Hierarchical Bayesian Formulations ................................................. 40

2.3.1 Stage 1: The Data Models ......................................................... 41

2.3.2 Stage 2: Process stage ............................................................... 41

2.3.3 Stage 3: Priors for the model and process stage parameters .......................... 45

2.3.4 Posterior distributions ..................................................................... 45

2.3.5 Simulating patterns from the two space-time models ............................... 48

2.4 Simulation example ........................................................................... 50

2.4.1 Simulation study for the marked space-time NHPPP model ........................ 50

2.4.2 Recovering the temporal process ...................................................... 51

2.4.3 Recovering the spatial (location) process .............................................. 52

2.4.4 Simulation study for the multivariate space-time model ......................... 53

2.4.5 Recovering the temporal process ...................................................... 54

2.4.6 Recovering the spatial (location) process .............................................. 54

2.5 Modeling Earthquake events in California, USA ............................................. 55

2.5.1 The Southern California Earthquake data .............................................. 55

2.5.2 Results based on the marked space-time NHPPP model ........................ 57

2.5.3 Results based on the multivariate space-time model ............................... 58

2.6 Discussion ..................................................................................... 59
Appendix: Full Conditional Distributions

2.A Full-Conditional Distributions

2.A.1 Updating the non-dynamic parameters

2.A.2 Updating the mean of number of components $\tau(t_k)$

2.A.3 Updating the mean of Dirichlet parameter $\mu_d(t_k)$

2.A.4 Updating the Dirichlet parameters $d(t_k)$

2.A.5 Updating the membership indicator vector $z_i(t_k)$

2.A.6 Updating the component probabilities $p(t_k)$

2.A.7 Updating the $j^{th}$ component mean $\mu_j(t_k)$

2.A.8 Updating the $j^{th}$ component covariance matrix $\Sigma_j(t_k)$

2.A.9 Updating the mean of the components means $\psi(t_k)$

2.A.10 Updating the covariance matrix $\Delta(t_k)$ of the component means $\psi(t_k)$

2.A.11 Updating the full-conditional distribution of $B(t_k)$

2.B Tables

2.C Figures

3 Spatial-Temporal Poisson Point Process Models via Conditioning on Location

3.1 Introduction

3.2 Modeling a spatial-temporal point process via conditioning on location

3.2.1 Modeling the event locations

3.2.2 Hawkes point processes

3.3 Hierarchical Bayesian Modeling

3.3.1 Stage 1: Data Model

3.3.2 Stage 2: Priors for the model parameters

3.3.3 Posterior distribution

3.3.4 Simulation of the space-time model conditioning on location

3.4 Simulated example
4.3.2 Modeling of the interaction effect ........................................ 164
4.3.3 Model distributions ....................................................... 164
4.3.4 Process models and prior distributions ................................. 166
4.3.5 Posterior sampling via double Metropolis-Hastings .................. 169

4.4 Simulating point patterns from the proposed MMSTPP model .......... 171

4.5 Simulation study ............................................................. 174
4.5.1 Simulations for model 1 .................................................. 174
4.5.2 Simulations for model 2 .................................................. 175
4.5.3 Simulations for model 3 .................................................. 175

4.6 Application to modeling tornado incidence ............................... 176
4.6.1 Results from model 1 ...................................................... 177
4.6.2 Results from model 2 ...................................................... 177
4.6.3 Results from model 3 ...................................................... 178

4.7 Discussion ................................................................. 178

Appendix: Full Conditional Distributions ........................................ 180
4.A Model 1: Posterior distribution and full conditionals ..................... 180
4.B Model 2: Posterior distribution and full conditionals ..................... 181
4.C Model 3: Posterior distribution and full conditionals ..................... 183
4.D Tables ........................................................................ 188
4.E Figures ........................................................................ 192

5 Future work and Conclusions ..................................................... 198
5.1 Extensions to the models for independent events ....................... 198
5.2 Extensions to the models for interacting events ......................... 199
5.3 Concluding remarks ......................................................... 200

Bibliography ........................................................................ 203
Vita ...................................................................................... 214
List of Figures

2.1 Model 1 (Simulation): Displaying the true Hawkes process intensity function (red), along with the estimated Hawkes process based on the posterior means (black) of the model parameters. We also present 95% credible sets for the Hawkes process based on the bounds of the 95% credible sets of the parameters. The temporal point pattern is shown at the bottom of the figure using red line segments. .................................................. 101

2.2 Model 1 (Simulation): Trace plots, histograms and ACF’s for $\mu$, $\alpha$ and $\beta$. The green dashed line represents the true value and the red line represents the posterior mean for the respective parameters. .............................. 102

2.3 Model 1 (Simulation): 2-d plots of the true Poison intensity surfaces (left column) and the corresponding PME intensity surfaces (right column) at times $t_1 = 2.6522$, $t_{13} = 25.3369$ and $t_{27} = 47.6197$. The point patterns are also shown and the gray crosses represent the mixture component means. ......................... 103

2.4 Model 1 (Simulation): 3-d plots of the true Poison intensity surfaces (left column) and the corresponding PME intensity surfaces (right column) at times $t_1 = 2.6522$, $t_{13} = 25.3369$ and $t_{27} = 47.6197$. .................................................. 104
2.5 Model 2 (Simulation): Displaying the true Hawkes process intensity function (red), along with the estimated Hawkes process based on the posterior means (black) of the model parameters. We also present 95% credible sets for the Hawkes process based on the bounds of the 95% credible sets of the parameters. The temporal point pattern is shown at the bottom of the figure using red line segments.

2.6 Model 2 (Simulation): Trace plots, histograms and ACF’s for $\mu$, $\alpha$ and $\beta$. The green dashed line represents the true value and the red line represents the posterior mean for the respective parameters.

2.7 Model 2 (Simulation): 2-d plots of the true Poison intensity surfaces (left column) and the corresponding PME intensity surfaces (right column) at times $t_1$, $t_{13}$ and $t_{27}$. The point patterns are also shown and the gray crosses represent the mixture component means.

2.8 Model 2 (Simulation): 3-d plots of the true Poison intensity surfaces (left column) and the corresponding PME intensity surfaces (right column) at times $t_1$, $t_{13}$ and $t_{27}$.

2.9 Earthquakes application: Displaying the time events of the Hawkes process. Note the spike in event arrivals as a result of the major earthquake at Baja, California, on April 4, 2010.

2.10 Model 1 (Earthquakes application): Displaying the estimated Hawkes process based on the posterior means of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments.

2.11 Model 1 (Earthquakes application): Trace plots, histograms and ACF’s for $\mu$, $\alpha$ and $\beta$. The red dashed line indicates the posterior means for the model parameters, $\mu$, $\alpha$ and $\beta$.

2.12 Model 1 (Earthquakes application): 2-d plots posterior means Poisson intensity surfaces at times $t_1$, $t_{26}$, $t_{27}$, $t_{28}$, $t_{29}$, and $t_{44}$. The point patterns are also shown and the yellow crosses represent the mixture component means.
2.13 Model 1 (Earthquakes application): 3-d plots of the true Poisson intensity surfaces (left column) and the corresponding PME Poisson intensity surfaces (right column) at times $t_1$, $t_{26}$, $t_{27}$, $t_{28}$, $t_{29}$, and $t_{44}$. ........................................ 112

2.14 Model 2 (Earthquakes application): Displaying the estimated Hawkes process based on the posterior means of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments. ............................ 113

2.15 Model 2 (Earthquakes application): Trace plots, histograms and ACF’s for $\mu$, $\alpha$ and $\beta$. The red dashed line indicates the posterior means for the model parameters, $\mu$, $\alpha$ and $\beta$. .......................................................... 114

2.16 Model 2 (Earthquakes application): 2-d plots PME Poisson intensity surfaces at times $t_1$, $t_{26}$, $t_{27}$, $t_{28}$, $t_{29}$, and $t_{44}$. The point patterns are also shown and the yellow crosses represent the component means. ........................................ 115

2.17 Model 2 (Earthquakes application): 3-d plots of the true Poisson intensity surfaces (left column) and the corresponding PME Poisson intensity surfaces (right column) at times $t_1$, $t_{26}$, $t_{27}$, $t_{28}$, $t_{29}$, and $t_{44}$. ........................................ 116

3.1 2-d plots of the true Poisson intensity surfaces (left) and the corresponding PME intensity surfaces (right). The point patterns are also shown and the gray crosses represent the mixture component means. ........................................ 136

3.2 3-d plots of the true Poisson intensity surfaces (left) and the corresponding PME intensity surfaces (right). ........................................ 136

3.3 Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=1. ........................................ 137

3.4 Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=1. ............ 138
3.5 Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=20. ........................................ 139

3.6 Displaying the estimated Hawkes process based on the posterior means (magenta), median (cyan), 95% credible intervals (lower-black,upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=20. ...................... 140

3.7 Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=40. ...................... 141

3.8 Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black,upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=40. ...................... 142

3.9 Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=60. ...................... 143

3.10 Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black,upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=60. ...................... 144

3.11 Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=80. ...................... 145

3.12 Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black,upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=80. ...................... 146

xii
3.13 Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=100. .............................................. 147

3.14 Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=100. .......... 148

3.15 Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=120. .............................................. 149

3.16 Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments location index=120. ............ 150

3.17 Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=140. .............................................. 151

3.18 Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=140. .......... 152

3.19 Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=160. .............................................. 153

3.20 Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=160. .......... 154
3.21 Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=180.

3.22 Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=180.

3.23 Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=202.

3.24 Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=202.

4.1 Model 1 Simulation (near hard-core model): displaying all events from time 1 to 4.

4.2 Model 1 Simulation (near hard-core model): displaying all events from time 5 to 10.

4.3 Model 2 simulation of a MSTPPP with random number of events at each time period. All the events are shown together in the top-left plot.

4.4 Model 2 simulation of a hard-core MSTPPP model with fixed number of events at each time period. All the events are shown together in the top-left plot. Note that within each time period, the events are further than $\rho = 1$ units away (hard-core distance).

4.5 Model 3 simulation with a fixed number of events at each time period. All the events are shown together in the top-left plot.
4.6 Tornadoes Application: displaying all tornado events from 2005 to 2010. The tornado power is indicated by a disc of different size.
List of Tables

2.1 Model 1 (Simulation): Summary statistics for the Hawkes process parameters. 81

2.2 Model 1 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=1 ($t_1 = 2.6522$). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 81

2.3 Model 1 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=13 ($t_{13} = 25.3369$). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82

2.4 Model 1 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=27 ($t_{27} = 47.6197$). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis). Note the index change from 1 to 2, and 2 to 1. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82

2.5 Model 2 (Simulation): Summary statistics for the Hawkes process parameters. 82

2.6 Model 2 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=1 ($t_1 = 2.6522$). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
2.7 Model 2 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=13 \((t_{13} = 25.3369)\). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis). Note the index change from 1 to 3, and 3 to 1.

2.8 Model 2 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=27 \((t_{27} = 47.6197)\). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis). Note the index change from 1 to 2, and 2 to 1.

2.9 Model 1 (Earthquakes application): Posterior estimates for the Hawkes process parameters.

2.10 Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=1.

2.11 Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=26.

2.12 Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=27.

2.13 Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=28.

2.14 Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=29.

2.15 Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=44.

2.16 Model 2 (Earthquakes application): Posterior estimates of the model parameters for the Hawkes process.

2.17 Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=1.
2.18 Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=26.

2.19 Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=27.

2.20 Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=28.

2.21 Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=29.

2.22 Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=44.

3.1 Summary of the true and estimated values of the mixture parameters. The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis).

3.2 Summary of spatial parameters.

3.3 Summary posterior Hawkes intensity function parameters at location index=1.

3.4 Summary posterior Hawkes intensity function parameters at location index=20.

3.5 Summary posterior Hawkes intensity function parameters at location index=40.

3.6 Summary posterior Hawkes intensity function parameters at location index=60.

3.7 Summary posterior Hawkes intensity function parameters at location index=80.

3.8 Summary posterior Hawkes intensity function parameters at location index=100.

3.9 Summary posterior Hawkes intensity function parameters at location index=120.

3.10 Summary posterior Hawkes intensity function parameters at location index=140.

3.11 Summary posterior Hawkes intensity function parameters at location index=160.

3.12 Summary posterior Hawkes intensity function parameters at location index=180.

3.13 Summary posterior Hawkes intensity function parameters at location index=200.
4.1 Model 1 (Near hard-core simulation, random number of events at each time period): Displaying the true values and Bayesian estimates of the model parameters. We run 20000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 50000 iterations of the Birth-Death algorithm. 188

4.2 Model 2 (Simulation, random number of events at each time period): Displaying the true values and Bayesian estimates for some of the model parameters. We run 20000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 50000 iterations of the Birth-Death algorithm. 188

4.3 Model 2 (Simulation of a hard-core model with fixed number of events at each time period): Displaying the true values and Bayesian estimates for some of the model parameters. We run 30000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 50000 iterations of the Birth-Death algorithm with change moves only. Note that $\xi$ is not affecting the number of points in this case ($n = 25$, fixed). 188

4.4 Model 3 (Simulation with fixed number of events at each time period): Displaying the true values and Bayesian estimates for some of the model parameters. We run 10000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 15000 iterations of the Birth-Death algorithm with change moves only. Note that $\xi$ is not affecting the number of points in this case ($n = 25$, fixed). 189

4.5 Tornadoes Application, Model 1: Displaying the Bayesian estimates of the model parameters along with 95% credible sets. We run 15000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 50000 iterations of the Birth-Death algorithm. 189
4.6 Tornadoes Application, Model 2: Displaying the Bayesian estimates of the model parameters along with 95% credible sets. We run 10000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 10000 iterations of the Birth-Death algorithm.

4.7 Tornadoes Application, Model 3: Displaying the Bayesian estimates of the model parameters along with 95% credible sets. We run 10000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 10000 iterations of the Birth-Death algorithm.
Dynamic Spatial-Temporal Point Process Models via Conditioning

Justin Obwoge Okenye

Abstract

We propose and investigate dynamic spatial-temporal point process models for independent and interacting events. The models for independent events are dynamic spatial-temporal Poisson point process (DSTPPP) model that account for temporal and spatial clustering. The models proposed for events with interaction are Markov (Gibbs) space-time point process models. We model the intensity function of a DSTPPP via conditioning arguments that allow for additional interpretations and inclusion of well-known point process models as special cases. Depending on the nature of the questions to be answered, the intensity function of a DSTPPP can be modeled in several ways. First, we develop models via conditioning on the time component of the events and then consider conditioning on the location component of the events. Modeling, simulation and computation are accomplished in a fully hierarchical Bayesian framework. Finally, we focus on dynamic marked Markov space-time point processes, where the events are allowed to interact with each other across time and space. Once again the hierarchical Bayesian framework is invaluable in this case, since it allows us to introduce dynamic process models via conditioning. The methodologies are illustrated using simulated data, and several applications, including modeling and inference for earthquake events in California, and tornado events in Missouri.
Chapter 1
Modeling Point Patterns

1.1 Introduction

Data in the form of a collection of points, randomly distributed in one, two, or three dimensions, arise in many scientific contexts. Examples of such processes include locations of trees in a forest, the longitude-latitude where a tornado first touches the ground, locations of nests of a colony of birds, forest fires, earthquakes, disease outbreaks, galaxies in the universe, or lightning strikes, and many more. In each of these examples we observe objects that are naturally described by their location and they are random in number. Often, some additional information (called a mark) is available that further describes the object, e.g., the dimensions of a tree, the type of bird, or the strength of a tornado or earthquake. We refer to such a dataset as a spatial point pattern and to the locations as its events. The goal of point process statistics is to analyze the geometrical structure of patterns formed by such objects over a window of observation $W$.

There are three possibilities for the locations of the events: they could be uniformly distributed over $W$, exhibit clustering, or exhibit regularity (i.e., equally spaced points over $W$). The choice of window introduces edge (boundary) effects that need to be considered in the modeling framework, since for a different choice of $W$, the statistical analysis may lead to different conclusions. In addition, the patterns could be informed by additional deterministic information (covariate information), such as a continuous field with values for any point in $W$, not just at the locations of the point pattern. Such important covariate information
may be included in the statistical model in order to provide a better understanding of the underlying stochastic mechanism that yields the locations of the events.

The random mechanism we use to model these point patterns is called a point process \( \mathcal{N} \), and a point pattern is treated as a realization of \( \mathcal{N} \). A random collection of points (or events) from a space \( \mathcal{X} \) is called a point process and the symbol \( \mathcal{N} \) has two meanings: (i) \( \mathcal{N} \) denotes a function operating on sets, and (ii) \( \mathcal{N} \) may also denote a random set, i.e., the set of points \( x_1, x_2, \ldots \), in the process. That is, \( \mathcal{N} = \{x_i\} \) or \( \mathcal{N} = \{x_1, x_2, \ldots\} \), with \( x_i \in \mathcal{X} \), and \( x \in \mathcal{N} \) means that the point \( x \in \mathcal{X} \) is in the set \( \mathcal{N} \). We may also write \( \mathcal{N} = \{x : x \in \mathcal{X}\} \) to denote the random collection of points \( \mathcal{N} \), and we use \( \mathcal{N}(\mathcal{W}) \) to denote the number of points of \( \mathcal{N} \) over a Borel set \( \mathcal{W} \subset \mathcal{X} \). Conditionally on \( n \), we will write \( \mathcal{N}(\omega) = \varphi_n = \{x_i\}_{i=1}^n \), to denote a realization of the point process \( \mathcal{N} \) (i.e., a point pattern), for some \( \omega \in \Omega \), with \( (\Omega, \mathcal{A}, P) \) a probability space.

When we observe the patterns over time, then we refer to the process as a space-time point process, and in the presence of marks, we say it is a marked space-time point process (MSTP-P). For foundations, modeling, applications, computation methods and evaluation of spatial, marked and space-time point processes we refer to the books by Ripley and Kelly (1977); Barndorff-Nielsen and Perez-Abreu (1999); Van Lieshout (2000); Moller and Waagepetersen (2003); Daley and Vere-Jones (2003); Diggle (2006); Illian et al. (2008); Gelfand et al. (2010); Cho et al. (2013); Lantuéjoul (2013); Chiu et al. (2013); Spodarev (2013); Cressie and Wikle (2015); Baddeley et al. (2015); Karr (2017); Adelfio and Chiodi (2020), and references therein).

An additional assumption that distinguishes point patterns is whether or not the locations of the events are independent from each other or if they interact in a systematic way. Under independence, the well known Poisson point process model is a natural starting point, whereas, for models with interactions, a Markov (or Gibbs) point process modeling approach is the usual way to proceed. Moreover, the type of statistical framework employed, classical or Bayesian, is another major difference between the types of models we can propose for point patterns.
Current contributions to the literature on (marked) space-time point processes either assume that the model parameters are the same across time or that the parameters depend on time but they are not dynamic, e.g., even though a model parameter \( \theta_t \) may depend on the time \( t \), there is no mechanism that evolves \( \theta_t \) into \( \theta_{t+1} \). For example, a recent contribution related to the methods explored here is King et al. (2012), where the authors illustrated (non-dynamic) Gibbs space-time models with mixed effects in order to model interacting locations of muskoxen herds. The authors utilized the area interaction process, a specific type of Gibbs model, along with elevation as covariate information, but their model was not dynamic so that they could not perform forecasting, and it required the assumption of conditional independence of locations across time given the covariate in order to be able to approximate the likelihood of the space-time point pattern. These are some of the difficulties one faces when working with such models, and as we see in later chapters, the methodology developed here eliminates these difficulties. In particular, this is where Bayesian hierarchical models (BHMs; see Section 2.3) are invaluable, since they allow us to introduce dynamic process models that make prediction of future states of the point process more accurate and efficient.

In this thesis both types of point process models are investigated in a hierarchical Bayesian framework. Novel marked Poisson space-time point process models are developed via conditioning arguments that allow for a natural interpretation of the different components of the intensity function, be it the location, the mark, or the time component. Moreover, novel dynamic Markov marked space-time point process models are developed, thus, extending the Gibbs point process models that where first considered as process models in the random set models of Micheas and Wikle (2009), and Dey and Micheas (2014).

Note that the rigorous definition for a point process \( \mathcal{N} \) and its distribution can be given in terms of the induced probability measure. Note that a measure \( \mu \) is locally finite if \( \mu(A) < \infty \), for all bounded sets \( A \in \mathcal{A} \), and \( \mu \) is called a counting measure if, for any finite \( A \in \mathcal{A} \), we have \( \mu(A) = card(A) = |A| = \# \) of points in \( A \), and if \( A \) is not finite assume \( \mu(A) = \infty \). Now let \( \mathcal{N} \) be a measurable map from a probability space \((\Omega, \mathcal{A}, P)\) to
the measurable space \((\mathcal{N}^f, \mathcal{N}^f)\), where \(\mathcal{N}^f\) denotes the collection of all locally finite counting measures on \(\mathcal{X}\), and \(\mathcal{N}^f\) denotes the generated \(\sigma\)-field from the collection of sets of the form \(\{\mathcal{N} \in \mathcal{N}^f : \mathcal{N}(B) = n\}\), for all \(n = 0, 1, 2, \ldots\), and \(B \in \mathcal{B}\), where \(\mathcal{B}\) the Borel sets of \(\mathcal{X}\). Letting \(\mathcal{N}_n^f = \{\mathcal{N} \in \mathcal{N}^f : \mathcal{N}(\mathcal{X}) = n\}\), we note that \(\mathcal{N}^f = \bigcup_{n=0}^{\infty} \mathcal{N}_n^f\). The distribution of \(\mathcal{N}\) is the induced probability measure given by

\[
Q_N(Y) = P(\mathcal{N}^{-1}(Y)) = P(\mathcal{N} \in Y) = P(\{\omega \in \Omega : \mathcal{N}(\omega) \in Y\}),
\]

for \(Y \in \mathcal{N}^f\).

In this introductory chapter the classical formulations of point process models that have appeared in the literature are given, along with relevant bibliography to the methodology described in subsequent chapters. Poisson point processes for independent events and their generalizations are briefly discussed in Section 1.2, whereas, Markov point process models and their generalizations are presented in Section 1.3. In Section 1.4, modeling of the conditional intensity function of a Poisson point process using mixture models is discussed, along with their modeling difficulties and how to overcome them. Finally, in Section 1.5 a summary of the thesis is provided.

### 1.2 Models for independent events

Consider a collection of \(n\) iid points in \(W^n\) i.e., let \(\varphi_n = \{x_i : x_i \in W, i = 1, 2, \ldots, n\}\) denote the point pattern. The simplest way to model this collection of points is by using the joint distribution of the sample

\[
f_n(x_1, x_2, \ldots, x_n|\theta) = f(x_1|\theta) \cdots f(x_n|\theta),
\]

conditionally on knowing \(n\), with \(\int_W f(x|\theta)dx = 1\). This is referred to as a standard multivariate analysis problem. When we introduce a distribution \(p(n)\) for \(n\), thereby making \(\mathcal{N}(W) = n\) random, then \(\varphi_n\) becomes a realization of a trivial point process \(\mathcal{N}\), which is
known as an iid cluster process. Therefore, the joint distribution of the point pattern $\varphi_n$ and $n$ is given by

$$f(\varphi_n, n|\theta) = f(x_1, x_2, \ldots, x_n, n|\theta) = f_n(x_1, x_2, \ldots, x_n)p(n),$$

and all inference on $\theta$ is based on this model. The iid cluster process with $p(n)$ being a Poisson distribution is a special case of the most commonly used point process model for independent events, the Poisson point process.

### 1.2.1 Poisson point processes

The most important point process for independent events is the Poisson process [Møller and Waagepetersen (2003); Daley and Vere-Jones (2003); Møller et al. (2006); Illian et al. (2008); Møller and Díaz-Avalos (2010); Adelfio and Chiodi (2020)]. It is often referred to as being “completely at random” or as a point process with “no interactions,” since the number of events (and the events themselves) over disjoint sets are independent of each other.

Consider a planar region $W \subset \mathcal{R}^2$ (extensions to higher dimensions are straightforward), and suppose that we observe $n$ points (events) $\{s_i\}_{i=1}^n$. In order to model this collection of points we consider the non-homogeneous Poisson point process (NHPPP) with intensity $\lambda(s)$, which assumes that the random variables (counting variables) $\mathcal{N}(B), B \subseteq W$, are Poisson, with mean measure

$$E[\mathcal{N}(B)] = \Lambda(B) = \int_B \lambda(s)ds,$$

where $\lambda(s)$ is known as the intensity function or surface for planar point patterns and uniquely determines the distribution of $\mathcal{N}$. In addition, the counts are independent over any finite collection of disjoint regions. The special case where $\lambda(s) = \lambda$ yields the homogeneous Poisson point process (HPPP), with rate $\lambda$, and mean measure $\Lambda(B) = \lambda |B|$, where $|B|$ denotes the area of $B$. 
The joint distribution of the independent events \( \varphi_n = \{s_i\}_{i=1}^n \), and the number of events \( N(W) = n \) over the window \( W \), is given by

\[
f(s_1, \ldots, s_n, n) = \frac{e^{-\Lambda(W)}}{n!} \prod_{i=1}^n \lambda(s_i) = \frac{1}{n!} e^{-\int_W \lambda(s) ds} \prod_{i=1}^n \lambda(s_i),
\]

for \( n \geq 0 \). Note that the intensity function uniquely determines the NHPPP, where \( \lambda(s)ds \) assumes the interpretation of the probability of finding a point in an infinitesimal ball of volume \( ds \) centered at \( s \).

In order to describe the point process, one needs to model its intensity \( \lambda(s) \), where the integral \( \int_W \lambda(s) ds \) is typically not available in closed form, requiring numerical approximation that introduces a computational burden. In order to achieve a better interpretation, as well as to simplify the calculation of this integral, we assume that

\[
\lambda(s|\theta) = \lambda \times \lambda_1(s|\theta_1),
\]

where \( \theta = (\lambda, \theta_1) \) are all the parameters of the intensity function, and \( \lambda_1 \) is a proper density over \( W \), so that

\[
\Lambda(W) = \int_W \lambda(s|\theta) ds = \lambda \int_W \lambda_1(s|\theta_1) ds = \lambda,
\]

and therefore \( \lambda > 0 \). This is interpreted as the average number \( \Lambda(W) = E[N(W)] \), of events over \( W \). Consequently, (1.1) becomes

\[
f(s_1, \ldots, s_n, n|\theta) = \frac{e^{-\lambda \lambda_1^n}}{n!} \prod_{i=1}^n \lambda_1(s_i|\theta_1),
\]

and we only need consider appropriate models for the proper density \( \lambda_1(s|\theta_1) \).

### 1.2.2 Marked Poisson point processes

Let \( \mathcal{N} \) be a Poisson point process on \( W \subseteq \mathbb{R}^2 \). Given some space \( \mathcal{M} \), if a random “mark” \( \xi \in \mathcal{M} \) is attached to each location \( s \in \mathcal{N} \), then \( \mathcal{N}_\xi = \{x = (s, \xi) : s \in \mathcal{N}, \xi \in \mathcal{M}\} \) is called a
marked NHPPP with locations in $\mathcal{W} \subseteq \mathcal{R}^2$ and mark space $\mathcal{M}$. For recent contributions in the literature on marked Poisson point processes see [Micheas 2011; Micheas et al. 2014; Zhou et al. 2014; Gold stick et al. 2015; Liu et al. 2020]

Consider observing a point pattern $\psi_n = \{x = (s_i, \xi_i) \in \mathcal{W} \times \mathcal{M}\}_{i=1}^n$, and assume that the events are independent of each other (no interaction). In order to model this collection of points, we consider a Poisson point process $\mathcal{N}$ over the domain $\mathcal{X} = \mathcal{W} \times \mathcal{M}$, which is characterized by its corresponding intensity function $\lambda(s, \xi)$. Then, the joint distribution of the point pattern $\psi_n$ and the number of events $n$ is given by

$$f(\psi_n, n) = \frac{e^{-\Lambda(\mathcal{W} \times \mathcal{M})}}{n!} \prod_{i=1}^n \lambda(s_i, \xi_i) = \frac{1}{n!} e^{-\int_{\mathcal{W} \times \mathcal{M}} \lambda(s, \xi) d\mathcal{s} d\xi} \prod_{i=1}^n \lambda(s_i, \xi_i), \quad (1.3)$$

where $\Lambda(\mathcal{W} \times \mathcal{M}) = \int_{\mathcal{W} \times \mathcal{M}} \lambda(s, \xi) d\mathcal{s} d\xi$. However, calculation of the integral requires approximations that lead to estimation procedures requiring numerical optimization and therefore, substantial computational burden. We alleviate this problem by modeling $\lambda(s, \xi)$ conditionally. There are two ways of modeling $\lambda(s, \xi)$ conditionally, since the event $x$ contains two types of information, the location and the mark components. Therefore, we can write

$$\lambda(s, \xi | \theta) = \lambda \times \lambda_1(\xi | \theta_1) \times \lambda_2[s | \xi, \theta_2(\xi)], \quad (1.4)$$

and treat this model as conditioning on the marks, or we can write

$$\lambda(s, \xi | \theta) = \lambda \times \lambda_1(s | \theta_3) \times \lambda_4[\xi | s, \theta_2(s)], \quad (1.5)$$

and think of this model as conditioning on location. In both cases, $\lambda > 0$ is some scaling parameter, $\lambda_1, \lambda_2, \lambda_3$ and $\lambda_4$ are proper densities, and

$$\lambda = \Lambda(\mathcal{W} \times \mathcal{M}) = \int_{\mathcal{W} \times \mathcal{M}} \lambda(s, \xi) d\mathcal{s} d\xi,$$

can be interpreted as the average number of events occurring in the window $\mathcal{W} \times \mathcal{M}$ (i.e.,
Considering the first kind of conditioning in (1.4), we can think of the point patterns created by this model as follows. For each mark $\xi_i$ there is a single event $s_i$ giving rise to a point pattern of the form
\[
\psi_n = \{(s_i(\xi_i), \xi_i), \ i = 1, \ldots, n\}. \tag{1.6}
\]
Now, using (1.3), the joint distribution of $\psi_n$ and $n$ is given by
\[
f(\psi_n, n|\theta) = \frac{e^{-\Lambda(W \times M)}}{n!} \prod_{i=1}^{n} \lambda(s(\xi_i), \xi_i|\theta) = \frac{e^{-\lambda n}}{n!} \prod_{i=1}^{n} \lambda_1(\xi_i|\theta_1) \lambda_2(s_i(\xi_i)|\xi_i, \theta_2(\xi_i)),
\]
where $\theta = (\lambda, \theta_1, \{\theta_2(\xi_i)\}_{i=1}^{n})$, denotes all the parameter of the model.

For the second kind of conditioning in (1.5), we can think of the point pattern being created as follows. For each event $s_i$ there is a single mark $\xi_i$ giving rise to a point pattern of the form
\[
\psi_n = \{((\xi(s_i), s_i), \ i = 1, \ldots, n\}, \tag{1.7}
\]
and therefore, using (1.3), the joint distribution of $\psi_n$ and $n$ can be written as
\[
f(\psi_n, n|\theta) = \frac{e^{-\Lambda(W \times M)}}{n!} \prod_{i=1}^{n} \lambda[\xi(s_i), s_i|\theta] = \frac{e^{-\lambda n}}{n!} \prod_{i=1}^{n} \lambda_1(s_i|\theta_1) \lambda_2[\xi(s_i)|\xi_i, \theta_2(s_i)],
\]
where $\theta = (\lambda, \theta_1, \{\theta_2(s_i)\}_{i=1}^{n})$, denotes all the parameters of the model.

### 1.2.3 Space-time Poisson point processes

We begin again by letting $\mathcal{N}$ be a Poisson point process on $W \subseteq \mathcal{R}^2$. A spatial-temporal point process $\mathcal{N}_t$ is an extension of $\mathcal{N}$ defined over some bounded region $W \times T$, where $W \subseteq \mathcal{R}^2$ and $T = [0, T) \subseteq \mathcal{R}^+$. We can represent $\mathcal{N}_t$ equivalently as a NHPPP by augmenting...
the space $W$ into the space $W \times T$, so that a point pattern from $\mathcal{N}_t$ is a collection of random points $\psi_n = \{x_1 = (s_1, t_1), \ldots, x_n = (s_n, t_n)\}$, random in number, where $\mathcal{N}_t(W \times T) = n$ represents the total number of points of the process $\mathcal{N}_t$ over the region $W \times T$. In this case, each event $x_i$ contains information that about the spatial location $s_i$ and the time $t_i$ associated with the $i-th$ event.

For recent contributions in the literature on space-time Poisson point processes (see Ogata, 1998; Møller and Waagepetersen, 2003; Diggle, 2006; Ji et al., 2009; Møller and Díaz-Avalos, 2010).

The standard assumption in space-time models (in particular, for independent locations) is that of separability. That is, the intensity of a space-time Poisson point process is expressed as $\lambda(s, t) = \lambda_1(s)\lambda_2(t)$, which can be restrictive and often an inappropriate assumption. Instead of imposing separability, we consider intensity functions via conditioning, similar in spirit to the marked spatial Poisson processes of Christou Micheas (2014) (note that these models did not consider a time component).

Now, in order to define the space-time NHPPP, we consider a Poisson process, which is characterized by the intensity function $\lambda(s, t)$, $s \in W$ and $t \in [0, T) \subseteq T$. One may choose different types of conditioning in describing $\lambda(s, t)$ depending on the nature of the problem. For example, when the data consist of a collection of points (locations) observed over a specific window at given times, then it makes sense to consider the conditioning approach on the time component. In contrast, if we have fixed locations that we visit over time (e.g., checking a certain spot in a forest for a specific species of bird), then conditioning on the spatial component is more appropriate.

The joint distribution of the point pattern $\psi_n$ and the number of events $n$ is given by

$$f(\psi_n, n) = \frac{e^{-\Lambda(W \times T)}}{n!} \prod_{i=1}^{n} \lambda(s_i, t_i) = \frac{1}{n!} e^{-\int_{t=0}^{T} \int_{W} \lambda(s, t) ds dt} \prod_{i=1}^{n} \lambda(s_i, t_i), \quad (1.8)$$

where $\Lambda(W \times T) = \int_{t=0}^{T} \int_{W} \lambda(s, t) ds dt$. 

9
In general, we can write the joint intensity either as:

\[
\lambda(s, t|\theta) = \lambda \times \lambda_1(t|\theta_1) \times \lambda_2[s|t, \theta_2(t)],
\]

(1.9)

treating this model as conditioning on the time component, or by

\[
\lambda(s, t|\theta) = \lambda \times \lambda_3(s|\theta_3) \times \lambda_4[t|s, \theta_2(s)],
\]

(1.10)

thinking this model as conditioning on location. In both cases, \(\lambda > 0\) is some scaling parameter, \(\lambda_1, \lambda_2, \lambda_3\) and \(\lambda_4\) are proper densities and

\[
\lambda = \Lambda(W \times T) = \int_{t=0}^{T} \int W \lambda(s, t)dsdt,
\]

can be interpreted as the average number of events occurring in the window \(T \times W\) (i.e., the average total number of events).

Based on the first type of conditioning of equation (1.9), we can think of the space-time patterns generated from \(\mathcal{N}_t\) as follows. For each temporal event \(t_i\) there is a single spatial event at location \(s_i\), and therefore, the resulting point patterns are of the form

\[
\psi_n = \{(s(t_i), t_i), i = 1, \cdots, n\}.
\]

(1.11)

Now using (1.8), the joint distribution of \(\psi_n\) and \(n\) is given by

\[
f(\psi_n, n|\theta) = \frac{e^{-\Lambda(W \times T)}}{n!} \prod_{i=1}^{n} \lambda(s(t_i), t_i|\theta)
\]

\[
= \frac{e^{-\lambda}}{n!} \prod_{i=1}^{n} \lambda_1(t_i|\theta_1) \lambda_2(s_i(t_i)|t_i, \theta_2(t_i)),
\]

where \(\theta = (\lambda, \theta_1, \{\theta_2(t_i)\})\), denotes all the parameters of the model.

When conditioning on location as in equation (1.10), we can think of the space-time patterns generated from \(\mathcal{N}_t\) as follows. For each spatial event \(s_i\) there is a single temporal
event \( t_i \) at that location, and therefore, the resulting point patterns are of the form

\[
\psi_n = \{(t(s_i), s_i), \, i = 1, \ldots, n\}.
\] (1.12)

Now using (1.8), the joint distribution of \( \psi_n \) and \( n \) is given by

\[
f(\psi_n, n|\theta) = \frac{e^{-\Lambda(W \times T)}}{n!} \prod_{i=1}^{n} \lambda(t(s_i), s_i|\theta)
\]

\[
= \frac{e^{-\lambda n}}{n!} \prod_{i=1}^{n} \lambda_1(s_i|\theta_1) \lambda_2(t_i(s_i)|s_i, \theta_2(s_i)),
\]

where \( \theta = (\lambda, \theta_1, \{\theta_2(s_i)\}_{i=1}^{n}) \), denotes all the parameters of the model.

### 1.2.4 Marked space-time Poisson point processes

The definition of the marked space-time NHPPP \( N_{\xi t} \) is straightforward. Since \( N_{\xi t} = \{x = (s, t, \xi) : s \in W, t \in T, \xi \in M\} \) consists of points \( x = (s, t, \xi) \), it can be thought of as a NHPPP over the augmented space \( \mathcal{X} = W \times T \times M \), where \( T \subset [0, +\infty) \), the time space, and \( M \) the mark space, so that the joint distribution is of the form (1.1), with the event \( s \) replaced by the augmented event \( x \).

For a review on recent advances in spatial-temporal point processes we refer to Greenspan (2013). For recent contributions in the literature on marked space-time Poisson point processes (see Ogata and Zhuang, 2006; Guan, 2008; Marsan and Lengline, 2008; Turner, 2009; Marsan and Lengline, 2010; Adelfio and Ogata, 2010; Mohler et al., 2011; Serra et al., 2014b; Fox et al., 2016; Hefley and Hooten, 2016; Raeisi et al., 2018).

Using \( \lambda(\cdot) \) to denote a generic intensity function, we can write the (joint) intensity \( \lambda(x) = \lambda(s, t, \xi) \), of \( N_{\xi t} \) via conditioning in several ways as follows:

\[
\lambda(x) = \lambda(s, t, \xi) = \lambda(s, t, \xi) \lambda(t, \xi) \lambda(s, \xi) = \lambda(s, t, \xi) \lambda(t, \xi) \lambda(s, \xi),
\] (1.13)

\[
\lambda(x) = \lambda(s, t, \xi) \lambda(t) = \lambda(s, t, \xi) \lambda(t) \lambda(s, t) = \lambda(s, t) \lambda(t) \lambda(s),
\] (1.14)

\[
\lambda(x) = \lambda(s, t, \xi) \lambda(s) = \lambda(s, t, \xi) \lambda(s) = \lambda(s) \lambda(t) \lambda(s).
\] (1.15)
where the third equalities (in parentheses) require conditional independence assumptions. The notation, $s_{\xi,t}$, simply means that we consider the locations for the given mark $\xi$ and $t$, while $\xi_t$ is used to denote the marks for the given time period $t$, and so forth.

One of the novelties of the methodology developed in this thesis is that this general framework allows for any of the well known models to be selected in formulating any of the corresponding conditional intensities. Note that even though it is natural to call the terms of the joint intensity “conditional intensities,” they are not the Papangelou conditional intensities (Papangelou, 1974). Furthermore, the conditioning proceeds as if we were dealing with densities, but in reality the intensities need not integrate to one. In particular, we assume that any conditional intensity term is analogous to a density, which we then model appropriately. For example, we set $\lambda(\xi_s|s) = \lambda_{\xi_s} \times f(\xi_s|s)$, where $\lambda_{\xi_s} > 0$, a constant, and $f(\xi_s|s)$ is some probability distribution on the mark space $\mathcal{M}$ that may depend on location $s \in \mathcal{W}$, and similarly for any other term.

The direction of conditioning that we select, i.e., model (1.13), (1.14), or (1.15), is determined by application and the type of question we want to answer. For example, on Sunday, May 22, 2011, Joplin Missouri, USA, was struck by a destructive tornado (F5 on the enhanced Fujita scale; for more details on the Fujita scale and its properties we refer to work of Wikle and Anderson (2003), resulting in over 150 deaths and $2.8$ billion in damages. This was the third tornado to strike Joplin since May, 1971. Clearly, obtaining the probability of observing an F2-5 tornado ($\xi = 1$) or F0-1 ($\xi = 0$) for the year $t = 2012$ and for Joplin’s spatial location $s^\ast$, would be of great importance in order to provide risk analysis for insurance and emergency management purposes.

This modeling can be accomplished via conditioning, and by considering the marked space-time NHPPP model with intensity given by (1.15). Then, in order to provide the probability of observing a destructive tornado, we simply require modeling and computation of $\lambda(\xi|s = s^\ast, t = 2012)$. Similar questions can be formulated in many scientific contexts, and we can model them using one of the three conditional marked space-time NHPPP models.
1.2.5 Why should we use conditioning?

The main advantage of the conditioning formulations above is that they provide a rich class of models that are easy to interpret and implement, since we can use existing models as special cases for parts of the joint intensity function. For example, if for a specific application we use the intensity $\lambda(x) = \lambda(s_\xi, t_\xi | \xi) \lambda(\xi)$, for a discrete mark taking values 0 or 1, then all we require is a model on the marks $\lambda(\xi) = \lambda_\xi f_\xi(\xi)$ (e.g., a multiple of a density $f_\xi(\xi)$ over the mark space), and for the term $\lambda(s_\xi, t_\xi | \xi)$, we can use any of the well known space-time models for the point process with events $(s_\xi, t_\xi)$ corresponding to mark $\xi$ (two space-time processes). For example, one might consider the two log-linear-trend surface models (Diggle, 2014) or the epidemic-type aftershock sequence (ETAS) model to model earthquake occurrences (Ogata, 2017), to name a few.

A general form for the joint intensity $\lambda(x)$ can lead to point patterns that are inappropriate, as well as violate the natural ordering of the time component. More precisely, classic models like the Ornstein-Uhlenbeck process can certainly be used to model the time component in the joint intensity. However, sampling from this process can yield points that do not adhere to the natural ordering of time. For example, the first observation could have a time period of 1.2 while the second is at time 0.5. This is not mathematically wrong, but not realistic and inappropriate for space-time data. For this reason, in Chapters 2 and 3 self-exciting processes are developed (Daley and Vere-Jones, 2003; Diggle, 2006; Ogata, 2017) that condition on the history of the process and adhere to the natural ordering of time. More importantly, BHMs, along with the conditioning of the intensity, lead to a full posterior distribution that is created based on independent chains that can be fit/simulated separately, as is shown in the chapters to follow.

1.3 Models for interacting events

In the previous section we discussed models for point processes in which we assumed that the events are independent, and in particular, we considered the Poisson point process
and its extensions. However, this approach will not work when the events are assumed to interact with each other, for example, when the point pattern shows signs of attraction or inhibition (repulsion). One way to introduce dependence between the events is by a Markov point process. One important class of such models is the family of Gibbs point processes known as pairwise interaction processes (Baddeley and Van Lieshout, 1995; Ripley, 2009). For recent developments (Van Lieshout, 2000; Ripley, 2009; Baddeley et al., 2015; Iftimi et al., 2015) we now define a Markov point process and its extensions.

1.3.1 Markov point process

In order to define a Markov point process we first define the concepts of a reflexive and symmetric relation, neighborhood, and clique. Assume that the space $(\mathcal{X}, \rho)$ is a Polish space (i.e., a complete separable metric space, with $\rho$ a metric).

Definition 1.1. (Reflexive and symmetric relation): A relation $\sim$ on $\mathcal{X}$ is said to be reflexive and symmetric if for any $x, y \in \mathcal{X}$, we have $x \sim x$ (reflexive) and $x \sim y \iff y \sim x$ (symmetric). The points $x, y \in \mathcal{X}$ are said to be neighbors if $x \sim y$. A standard choice for a reflexive and symmetric relation $\sim_r$ on $\mathcal{X}$ is obtained by $x \sim_r y \iff \|y - x\| < r$, for some fixed $r > 0$.

Definition 1.2. (Neighborhood and Cliques): A neighborhood of a set $A \subseteq \mathcal{X}$ denoted by $\partial(A)$ is defined as $\partial(A) = \{x \in \mathcal{X} : x \sim y \text{ for some } y \in A\}$. If $A = \{y\}$, a singleton, then $\partial(\{y\}) = \{x \in \mathcal{X} : x \sim y\}$, is the neighborhood of the point $y$. A clique is a subset of $\mathcal{X}$ with the property that all of its points are neighbors of each other with respect to some relation $\sim$.

Based on these basic definitions, we are now able to define a Markov point process $\mathcal{N}$ with distribution $Q_{\mathcal{N}}$. However, we first note that the Poisson point process distribution $\Pi$ (typically HPPP with intensity $\lambda = 1$) can be used to define the density $p_{\mathcal{N}}$ of any finite point process as the Radon-Nikodym derivative $[\frac{dQ_{\mathcal{N}}}{d\Pi}]$, and therefore $\Pi$ assumes a
similar role as that of the Lebesgue or counting measure in the definition of a (continuous or
discrete) random variable. For an excellent exposition of Markov point processes and such
constructions we refer to [Van Lieshout, 2000].

**Definition 1.3. (Markov Point Process):** Let \( p_N : \mathbb{N}^f \to [0, \infty) \) be a function such
that for all \( \varphi \in \mathbb{N}^f \) with \( p_N(\varphi) > 0 \), we have

(i) \( p_N(\psi) > 0 \) for all \( \psi \subseteq \varphi \), and

(ii) for all \( x \in \mathcal{X} \setminus \varphi \), the ratio \( p_N(\varphi \cup \{x\})/p_N(\varphi) \), depends only on \( x \) and the neighbors
of \( x \) from \( \varphi \); that is, the points in the neighborhood \( \partial(\{x\}) \cap \varphi = \{y \in \varphi : x \sim y\} \).

Then, \( p_N \) is called a **Markov function** with respect to a reflexive and symmetric relation
\( \sim \) and a point process with density \( p_N \) with respect to a Poisson point process (typically of
unit intensity) is called a **Markov point process** on \( \mathcal{X} \).

For a simple process (distinct points) \( \mathcal{N} \) defined on a Polish space \((\mathcal{X}, \rho)\) equipped with
a Borel measure \( \mu \) (typically Lebesgue), the Papangelou conditional intensity ([Papangelou,
1974]) is defined as the measurable function \( \lambda(x|\mathcal{N}) : \mathcal{X} \times \mathbb{N}^f \to \mathcal{R}_0^+ \), that satisfies

\[
E^\mathcal{N}\left( \sum_{x \in \mathcal{N}} g(x, \mathcal{N} \setminus \{x\}) \right) = \int \int g(x, \varphi)\lambda(x|\varphi)d\mu(x)Q_\mathcal{N}(d\varphi),
\]

(1.16)

for any measurable function \( g(x, \mathcal{N}) : \mathcal{X} \times \mathbb{N}^f \to \mathcal{R}_0^+ \), where \( E^\mathcal{N} \) denotes expectation
with respect to the distribution \( Q_\mathcal{N} \) of the point process \( \mathcal{N} \). Under certain conditions (see
[Van Lieshout, 2000, pg 41]), it can be shown that for \( x \not\in \varphi \), we have

\[
\lambda(x|\varphi) = p_N(\varphi \cup \{x\})/p_N(\varphi),
\]

(1.17)

and therefore, in order to define Markov point process models, we need to work with the
Papangelou conditional intensity. In particular, given \( \lambda(x|\varphi) \), where \( \varphi = \{x_1, \ldots, x_n\} \) a
point pattern, then the density \( p_N \) with respect to \( \mu \) is obtained via (1.17) by

\[
p_N(\varphi) \propto \prod_{i=1}^n \lambda(x_i|\{x_1, \ldots, x_{i-1}\}).
\]

(1.18)
We can think of $\lambda(x|\varphi)d\mu(x)$ as the conditional probability that there is a point of $\mathcal{N}$ in an infinitesimal sphere $b(x)$ of volume $d\mu(x)$ containing $x$, given the realization $\varphi$ of $\mathcal{N}$ outside of $b(x)$. In order to write the joint density of a Markov point process $p_{\mathcal{N}}(\varphi)$, we appeal to the famous Hammersley-Clifford theorem.

**Theorem 1.1. (Hammersley-Clifford):** A point process density $p_{\mathcal{N}}: \mathbb{N}^f \to [0, \infty)$ of a finite point process $\mathcal{N}$ is Markov with respect to a neighborhood relation $\sim$ iff there exists a measurable function $p_C: \mathbb{N}^f \to [0, \infty)$ such that

$$p_{\mathcal{N}}(\varphi) = \prod_{\text{all cliques } \psi \subseteq \varphi} p_C(\psi), \quad (1.19)$$

for all finite point patterns $\varphi \in \mathbb{N}^f$.

For a proof in the context of point processes (Van Lieshout, 2000, pg 49). The clique function $p_C$ operates on all cliques $\psi$ of a point pattern $\varphi_n = \{x_1, x_2, \ldots, x_n\}$, and can be thought of as the local characteristics of $\mathcal{N}$. It plays the role of the full conditional distributions of a Markov random field (MRF; Besag (1974)). Since the cliques could be of any size, the function $p_C$ operates on cliques of size 0 (empty set), 1, 2, 3, and so forth. The Hammersley-Clifford theorem is useful in breaking up a high-dimensional joint distribution $p_{\mathcal{N}}(\varphi)$ into manageable clique interaction functions $p_C(\psi)$ that are easier to interpret and have a lower dimension. However, one has to verify that a particular choice of interaction function $p_C(\psi)$ results in a $p_{\mathcal{N}}(\varphi)$ that is integrable.

Gibbs point processes were originally introduced in physics by Bryan (1902). When we define them as finite point process models, we assume that we study $\mathcal{N}$ over some bounded set $\mathcal{X}$, with $\mathcal{N}(\mathcal{X}) = n$, and then define the joint density for the number of events and their locations using the density

$$f(\varphi_n, n) = c^{-1}\left[e^{-\mu(\mathcal{X})}/n!\right]\exp\{-En(\varphi_n)\} = c^{-1}\left\{e^{-\mu(\mathcal{X})}/n!\right\}\exp\left\{-\sum_{\psi \subseteq \varphi_n, \psi \neq \emptyset} V(\psi)\right\}, \quad (1.20)$$

with respect to the Poisson process with unit intensity, where $c$ is the (usually intractable)
normalizing constant known as the partition function, \( E_n(\varphi_n) = \sum_{\emptyset \neq \psi \subseteq \varphi_n} V(\psi) \) denotes the potential energy function, and \( V(\psi) \) the potential of \( \psi \), with \( \psi \) any non-empty subset of \( \varphi_n \). Adding a nearest-neighbor condition, for example, \( V(\psi) = 0 \), for all subsets \( \psi \) with more than two points, turns the Gibbs process into a Markov process (see Cressie and Wikle, 2015, pg.675).

We can rewrite equation (1.20) equivalently in terms of \( k \)-level (clique) interaction functions \( g_1 \ldots k(\cdot) \) as

\[
f(\varphi_n, n) = c^{-1} \left[ e^{-\mu(X)/n!} \right] \exp \left\{ \sum_{i=1}^{n} g_1(x_i) + \sum_{1 \leq i < j \leq n} g_{12}(x_i, x_j) + \cdots + g_{1 \ldots n}(x_1 \ldots x_n) \right\} ,
\]

(1.21)

\( x_i \in \mathcal{X} \), \( i = 1, 2, \ldots, n \), \( n = 1, 2, \ldots \), where \( c \) the normalizing constant and \( N(\mathcal{X}) = n \), for some bounded set \( \mathcal{X} \). Choosing different forms for the functions \( g_{1 \ldots k}(\cdot) \), leads to different Markov point process models that may require additional conditions in order for (1.21) to be integrable. In particular, the Markov point process is said to be stable if \( 0 < c < \infty \). A sufficient condition is that the potential energy in (1.20) is bounded, namely, \( E_n(\varphi_n) \geq -an \), for all \( n \geq 0 \), and some constant \( 0 \leq a < \infty \) (Ruelle, 1999).

The simplest example of Markov point processes is the inhomogeneous pairwise interaction Markov point process. Let \( \varphi_n = \{x_1, x_2, \ldots, x_n\} \) denote a point pattern, where \( N(\mathcal{W}) = n \), and let \( \sim \) some reflexive and symmetric relation that is used to describe neighbors and cliques. The inhomogeneous pairwise interaction Markov point process is obtained by considering cliques of size 1 and 2, with the density given by

\[
p_N(\varphi_n) = a \exp \left\{ \sum_{i=1}^{n} b(x_i) + \sum_{1 \leq i < j \leq n} c(x_i, x_j) \right\} ,
\]

where \( a > 0 \), the normalizing constant, \( b : \mathcal{W} \to [0, \infty) \) the intensity function and \( c : \mathcal{W} \times \mathcal{W} \to [0, \infty) \) the pair interaction function. The function \( b \) can be thought of as modeling "large-scale effects", whereas, the function \( c \) models 'small-scale effects'. Note that this model is also known as the pair-potential process in statistical mechanics. The process is called
inhomogeneous since $b(x)$ is a general function that depends on $x$. Clearly, $p_N$ is of the form expressed in equation (1.19), with clique interaction function $p_C$ defined as: $p_C(\emptyset) = a$, $p_C(\{x\}) = b(x), x \in W$, $p_C(\{x_i, x_j\}) = c(x_i, x_j), x_i, x_j \in W$, and for cliques of size 3 and above $p_C(\cdot) = 0.$

1.3.2 Marked Markov point process

The definition of a marked Markov point process readily extends from definition 3 of a general Markov point process, by augmenting the original space using a mark space $M$. Let $N_f^\xi$ the collection of all locally finite counting measures on $X = W \times M$.

Definition 1.4. (Marked Markov Point Process): A function $p_{N_\xi}: N_f^\xi \to [0, \infty)$ is called Markov with respect to a reflexive and symmetric relation $\sim$ on $W \times M$, if for all $\varphi \in N_f^\xi$ with $p_{N_\xi}(\varphi) > 0$, we have:

(i) $p_{N_\xi}(\psi) > 0$ for all $\psi \subseteq \varphi$, and

(ii) for any $x = (s, \xi) \in W \times M$, the ratio $p_{N_\xi}(\varphi \cup \{(s, \xi)\})/p_{N_\xi}(\varphi)$, depends only on $(s, \xi)$ and the neighbors of $(s, \xi)$ from $\varphi$, that is, the points in the neighborhood $\partial(\{(s, \xi)\}) \cup \varphi = \{(u, \zeta) \in \varphi : (s, \xi) \sim (u, \zeta)\}$.

A marked Markov point process on $W \times M$ is a finite point process with density $p_{N_\xi}$, such that $p_{N_\xi}$ is a Markov function.

In order to write the joint density of a Markov point process $p_{N_\xi}(\varphi)$, we once again appeal to the Hammersley-Clifford theorem. A typical choice for a reflexive and symmetric relation on $W \times M$ denoted by $\sim_m$ assumes independence between the marks is given by $x = (s, \xi) \sim_m y = (u, \zeta) \iff \|s - u\| < r$, for some fixed $r > 0$.

The simplest example of a marked Markov point processes is the inhomogeneous pairwise interaction marked Markov point process. Let $\varphi_n = \{x_i = (s_i, \xi_i), s_i \in W, \xi \in M, i = 1, 2, \ldots, n\}$ denote a marked point pattern, where $N_\xi(W \times M) = n$, and let $\sim$ some reflexive and symmetric relation that is used to describe neighbors and cliques. The inhomogeneous pairwise interaction marked Markov point process is obtained by considering cliques of size
1 and 2, with the density given by

\[ p_{N_{\xi}}(\varphi_n) = a \exp \left\{ \sum_{i=1}^{n} b(s_i, \xi_i) + \sum_{1 \leq i < j \leq n} c(s_i, \xi_i, s_j, \xi_j) \right\}, \tag{1.22} \]

where \( a > 0 \), the normalizing constant, \( b : \mathcal{W} \times \mathcal{M} \to [0, \infty) \) the intensity function and \( c : (\mathcal{W} \times \mathcal{M}) \times (\mathcal{W} \times \mathcal{M}) \to [0, \infty) \) the pair interaction function. Clearly, \( p_{N_{\xi}} \) is of the form of Equation (1.19), with clique interaction function \( p_C \) defined as:

\[ p_{C}(\emptyset) = a, \quad p_{C}(\{x\}) = b(x), \quad x \in \mathcal{W} \times \mathcal{M}, \quad p_{C}(\{x_i, x_j\}) = c(x_i, x_j), \quad x_i, x_j \in \mathcal{W} \times \mathcal{M}, \]

and for cliques of size 3 and above \( p_C(\cdot) = 0 \).

### 1.3.3 Space-time Markov point process

Based on Definition 3, the definition of a space-time Markov point process is straightforward by once again augmenting the original space to \( \mathcal{W}_s \times \mathcal{W}_t \). Therefore, a space-time Markov point process is simply a Markov point process defined on \( \mathcal{X} = \mathcal{W}_s \times \mathcal{W}_t \subseteq \mathbb{R}^2 \times \mathbb{R}^+ \).

Let \( \mathbb{N}_{st}^f \) the collection of all locally finite counting measures on \( \mathcal{X} \).

**Definition 1.5. (Space-Time Markov Point Process):** A function \( p_{N_{st}} : \mathbb{N}_{st}^f \to [0, \infty) \) is called Markov with respect to a reflexive and symmetric relation ~ on \( \mathcal{W}_s \times \mathcal{W}_t \), if for all \( \varphi \in \mathbb{N}_{st}^f \) with \( p_{N_{st}}(\varphi) > 0 \), we have:

(i) \( p_{N_{st}}(\psi) > 0 \) for all \( \psi \subseteq \varphi \), and

(ii) for any \( (s, t) \in \mathcal{W}_s \times \mathcal{W}_t \), the ratio \( p_{N_{st}}(\varphi \cup \{(s, t)\}) / p_{N_{st}}(\varphi) \), depends only on \( (s, t) \) and the neighbors of \( (x, t) \) from \( \varphi \); that is, the points in the neighborhood \( \partial(\{(s, t)\}) \cup \varphi = \{(u, v) \in \varphi : (s, t) \sim (u, v)\} \).

A space-time Markov point process on \( \mathcal{W}_s \times \mathcal{W}_t \) is a finite point process with density \( p_{N_{st}} \), such that \( p_{N_{st}} \) is a Markov function.

In order to write the joint density of a space-time Markov point process \( p_{N_{st}}(\varphi) \), we once again appeal to the Hammersley-Clifford theorem. A typical choice for the symmetric and reflexive relation is given by \( x_i \sim_{st} x_j \Leftrightarrow ||s_i - s_j|| \leq r \), and \( t_j = t_i + 1 \) (future) or \( t_j = t_i \) (past).
(present) or \( t_j = t_i - 1 \) (past), where \( \mathbf{x}_i = (s_i, t_i), \ i = 1, 2, \ldots, n. \)

The simplest example of a space-time Markov point processes is the inhomogeneous pairwise interaction Markov process. Let \( \varphi_n = \{ \mathbf{x}_i = (s_i, t_i), \ s_i \in \mathcal{W}_s, \ t_i \in \mathcal{W}_t, i = 1, 2, \ldots, n \} \) denote a space-time point pattern, where \( \mathcal{N}_{st}(\mathcal{W}_s \times \mathcal{W}_t) = n \), and let \( \sim \) some reflexive and symmetric relation that is used to describe neighbors and cliques. The inhomogeneous pairwise interaction space-time Markov process is obtained by considering cliques of size 1 and 2, with the density given by

\[
p_{N_{st}}(\varphi_n) = a \exp \left\{ \sum_{i=1}^{n} b(s_i, t_i) + \sum_{1 \leq i < j \leq n} c(s_i, t_i, s_j, t_j) \right\},
\]

where \( a > 0 \), the normalizing constant, \( b : \mathcal{W}_s \times \mathcal{W}_t \rightarrow [0, \infty) \) is the intensity function and \( c : (\mathcal{W}_s \times \mathcal{W}_t) \times (\mathcal{W}_s \times \mathcal{W}_t) \rightarrow [0, \infty) \) denotes the pair interaction function. Clearly, \( p_N \) is of the form expressed in Equation (1.19), with clique interaction function \( p_C \) defined as:

\[
p_C(\emptyset) = a, \ p_C(\{ \mathbf{x} \}) = b(\mathbf{x}), \ \mathbf{x} \in \mathcal{W}_s \times \mathcal{W}_t, \ p_C(\{ \mathbf{x}_i, \mathbf{x}_j \}) = c(\mathbf{x}_i, \mathbf{x}_j), \ \mathbf{x}_i, \mathbf{x}_j \in \mathcal{W}_s \times \mathcal{W}_t, \text{ and for cliques of size 3 and above } p_C(\cdots) = 0. \]

We extend these models to the marked space-time Markov point process case in chapter 4 (see section 4.2), where we study specific models using a hierarchical Bayesian framework.

### 1.4 Mixture models

Spatial point patterns are typically highly heterogeneous and require flexible models for the underlying intensity functions. To provide flexibility in characterizing spatial heterogeneity, we consider modeling the intensity function by employing a finite mixture framework in which the intensity function \( \lambda(\mathbf{s}) \) is a product of some scaling parameter \( \lambda > 0 \) and a density of a mixture of bivariate normal distributions.

Modeling the intensity function of a point process can be accomplished in both parametric and non-parametric Bayesian settings. In the parametric Bayesian setting, the prior beliefs are formally defined as probability distributions for parameters of an assumed model, whereas
in the non-parametric Bayesian setting the assumption of a specific parametric model is
dropped and instead, “parameters” correspond to the unknown distribution of the observed
data and the prior beliefs which are defined as probability measures on the distributions that
may have generated the data (e.g. Sethuraman, 1994).

The non-parametric approach has been used successfully to model spatial point processes,
in particular, using Dirichlet process (DP) models or infinite mixtures (Escobar and West,
1995; Wolpert and Ickstadt, 1998; Scricciolo, 2006; Kottas and Sansó, 2007; Ji et al., 2009)
approaches have also been used to model spatio-temporal point processes (Kottas and Sansó,
2007; Taddy, 2008; Ji et al., 2009; Ding et al., 2012; Kottas et al., 2012). Non-parametric
methods are quite flexible and can describe densities or intensity functions of point processes
well; however, they are highly complex mathematically and as such present computational
challenges.

The parametric approach has also been used to model spatial point processes, particularly
using a finite mixtures framework. Despite of the fact that there is a plethora of papers in the
literature on finite mixtures, this approach has not been very popular due to the intractable
form of the likelihood especially when the number of components is unknown. Dempster
et al. (1977) introduced the concept of data augmentation alongside the EM algorithm,
which made the likelihood of finite mixture models more tractable and the estimation of
the parameters of finite mixtures feasible. Furthermore, the Gibbs samplers by Tanner and
Wong (1987) and Diebolt and Robert (1994) constitute significant developments towards
building a hierarchical Bayesian framework that helped with Bayesian estimation for the
parameters of a finite mixture with a fixed number of components.

The finite mixture approach is particularly useful when specific features of the intensity
function are of interest, since they can provide inference for individual features of an intensity
surface. For example, Chakraborty and Gelfand (2010) considered a finite Gaussian mixture
for a small and fixed number of components for the intensity function of a NHPPP. This
idea was further extended by Christou Micheas (2014) to include marked NHPPPs and, more
importantly, for the case of varying dimension finite mixtures.
1.4.1 Modeling the intensity function as a mixture

Suppose that we observe \( n \) the spatial events \( \varphi_n = \{s_i, \ i = 1, \ldots, n\} \), where \( s_i \) represents the spatial location of the \( i \)–\( th \) event and further assume that the events are independent of each other (no interaction). In order to model this collection of points we consider a Poisson point process \( \mathcal{N} \) over the domain \( \mathcal{W} \), with joint distribution given by (1.2), and consider modeling \( \lambda_1(s|\theta_1) \) using an \( m \)-component mixture of bivariate normal components by

\[
\lambda_1(s|\theta_1) = \sum_{j=1}^{m} p_j \varphi_j(s|\mu_j, \Sigma_j), \tag{1.24}
\]

where \( m \) denotes the number of components, \( \mu_j \) the \( j \)–\( th \) component mean vector, \( \Sigma_j \) the \( j \)–\( th \) component covariance matrix, \( p_j \) the \( j \)–\( th \) component probability. Further, \( \varphi_j(s|\mu_j, \Sigma_j) \) denotes the density of a \( N_2(\mu_j, \Sigma_j) \) and \( \theta_1 = \{p_j, \mu_j, \Sigma_j\}_{j=1}^{m} \) denotes all the parameters of the \( m \)-component mixture of bivariate normal components. When a component has positive probability outside \( \mathcal{W} \) (edge effects) then the component can be truncated to integrate to one by letting

\[
\Phi_j(\mu_j, \Sigma_j) = P(s \in \mathcal{W}|\mu_j, \Sigma_j) = \int_{\mathcal{W}} \varphi_j(s|\mu_j, \Sigma_j)ds, \tag{1.25}
\]

and then replacing \( \varphi_j(s|\mu_j, \Sigma_j) \) in Equation (1.24) with \( \frac{\varphi_j(s|\mu_j, \Sigma_j)}{\Phi_j(\mu_j, \Sigma_j)} \).

Now, using the parametric model of Equation (1.24), the likelihood based on \( n \) observed events in the point pattern \( \varphi_n \) can be written as

\[
L(\theta|s_1, \ldots, s_n, n) = \frac{e^{-\lambda} \lambda^n}{n!} \prod_{i=1}^{n} \sum_{j=1}^{m} p_j \varphi_j(s|\mu_j, \Sigma_j), \tag{1.26}
\]

\( n \geq 0 \), which is intractable. In order to alleviate the intractability in (1.26) we use the hierarchical Bayesian formulation based on data augmentation by Christou Micheas (2014).
1.4.2 Data augmentation

Following [Dempster et al. (1977)], any mixture model can be thought of as a missing or incomplete data model, where for each observed \( s_i \) we define a vector of latent variables \( z_i = [z_{i,1}, \ldots, z_{i,m}]^T \) which indicates to which component \( s_i \) belongs (membership indicator). More precisely, \( z_{i,j}(t) = 1 \) if the \( i-th \) comes from the \( j \)-th component, or 0 otherwise, with \( \sum_{j=1}^{m} z_{i,j} = 1 \), for each \( i = 1, \ldots, n \). Note that \( p(z_{i,j} = 1|\theta_1) = p_j \) and \( f(s_i|z_{i,j} = 1, \theta_1) = \varphi_j(s_i|\mu_j, \Sigma_j) \), for \( i = 1, \ldots, n \) and \( j = 1, \ldots, m \). In the point process context we can think of \( z_{i,j} \) as a mark associated with \( s_i \) so that \( z_i = [z_{i,1}, \ldots, z_{i,m}]^T \in \mathcal{M}_1 = \{[z_1, \ldots, z_m]: z_{i,j} \in \{0,1\} \text{ and } \sum_{j=1}^{m} z_{i,j} = 1\} \), where \( \mathcal{M}_1 \) is the mark space. Augmenting each \( s_i \) by \( z_i \) we obtain the “completed point intensity function” given by

\[
\lambda_1(s, z_1, \theta_1) = \prod_{j=1}^{m} p_j z_{i,j}^{z_{i,j}} \left\{ \varphi_j(s_i|\mu_j, \Sigma_j) \right\}^{z_{i,j}},
\]

(1.27)

As a result, the likelihood of Equation (1.26), based on \( \varphi_n = \{s_i, \ i = 1, \ldots, n\} \) becomes

\[
L(\theta|s_1, \ldots, s_n, n) = \frac{e^{-\lambda} \lambda^n}{n!} \prod_{i=1}^{n} \prod_{j=1}^{m} p_j z_{i,j}^{z_{i,j}} \left\{ \varphi_j(s_i|\mu_j, \Sigma_j) \right\}^{z_{i,j}},
\]

(1.28)

and all inference is based on this likelihood.

1.5 Thesis outline

The remainder of the thesis is organized as follows. In Chapter 2, the objective is to model the intensity function of a space-time point process dynamically. In particular, space-time marked non-homogeneous Poison point processes (STMNHPPP) models and the dynamic space-time non-homogeneous Poison point processes (DSTNHPPP) models via conditioning on the time component are developed. Two modeling approaches are proposed; namely, We propose a STMNHPPP that can be viewed as a multi-type point process and a multivariate space-time model that models the point pattern directly assuming that the number of
points is fixed. Note that both approaches are capable of handling replications in the time component. Furthermore, modeling via conditioning helps to alleviate the problem of having to approximate the intractable integrals involved, as well as, helping to achieve a natural interpretation for the model parameters.

In Chapter 3 dynamic STMNHPPP models via conditioning on the location component of the event are considered. Since, in this case, all parameters depend on the location component for all \( s \in \mathcal{W} \), we define and work with random fields in order to model the parameters dynamically. Furthermore, the hierarchical Bayesian framework adopted in Chapters 2 and 3 make the models amenable to the inclusion of covariate information by treating such information as marks or including it at stage 2, or utilize important \textit{a priori} information in the creation of the priors and hyper-priors at stages 2 and 3 (see section 2.3).

Chapter 4 focuses on modeling Markov space-time marked point processes. The hierarchical Bayesian framework is invaluable in this case, since it allows us to introduce dynamic process models via conditioning. Introducing and illustrating all the details of dynamic process models for the parameters of the point process is one of the major contributions of this chapter. In particular, three approaches to modeling the main effect are considered and a single approach to modeling the interaction effect is used. Finally, a summary of the dissertation, as well as possible extensions and future work, is discussed in Chapter 5.
Chapter 2

Dynamic Spatio-Temporal Poisson Point Process Models via Conditioning on Time

2.1 Introduction

Major earthquakes are known to increase the risk of subsequent earthquakes (aftershocks) near their location (around the epicenter) and shortly (hours/days/weeks) after their occurrence; such a behavior is often referred to as space-time clustering. In order to uncover the underlying spatial-temporal structure, space-time clustering is modeled using a spatio-temporal point process (STPP). Spatial and spatial-temporal point process models have emerged as powerful tools that help model the underlying structure of events in space as they evolve over time.

In recent years, the point process theory and applications literature has documented both the spatial (e.g. Diggle 2006, Mohler et al. 2011, Baddeley et al. 2015, Cressie and Wikle 2015), and the temporal cases (e.g. Cox and Isham 1980, Daley and Vere-Jones 2003, Lee and Wilkinson 2020). However, theory on the analysis of spatio-temporal data is less well established; see for example Diggle (2006) and Cressie and Wikle (2015).
The realization of a point process model is known as a point pattern and it consists of points (otherwise known as events) that may contain more information than just the location and the time component of the event, namely, a mark value. Marks can be used to describe additional information about the event, for example, the magnitude of an earthquake event on the Richter scale. Such point process models are called Marked point processes (see, for example, Illian et al. (2008).

Both spatial and spatial-temporal point patterns often exhibit clustering and thus, the patterns are characterized by a set of parameters that are used to describe each cluster. Much of the theory of spatial-temporal point processes carries over from that of spatial point processes. However, in modeling spatial-temporal processes, we need to account for the natural ordering of time, which is not considered when modeling purely spatial processes. Thus, it is often convenient to view a spatial-temporal point process as a marked temporal point process, where each time point is associated with a spatial location (Schoenberg et al. (2002). In this chapter, we consider spatial point patterns that evolve over time in the sense that the locations of the clusters, number of clusters, and number of points are time dependent (i.e.; the spatial point patterns consist of clusters whose number and locations are time-varying). Modeling DSTPPs involves modeling the intensity function of the process, which then is utilized in the formulation of the corresponding likelihood function for the point pattern. Standard models include the non-homogeneous Poisson point process (NHPPP) for independent locations, Gibbs processes for interacting locations, and their respective generalizations.

Modeling the intensity function of a point process can be accomplished in both parametric and non-parametric Bayesian settings. In the parametric Bayesian setting, prior beliefs are formally defined as probability distributions for parameters of an assumed model, whereas in the non-parametric Bayesian setting, the assumption of a specific parametric model is dropped and instead, “parameters” correspond to the unknown distribution of the observed data and the prior beliefs are defined as probability measures on the distributions that may have generated the data (Sethuraman 1994).
The non-parametric approach has been used successfully to model spatial point processes, in particular, using Dirichlet process (DP) models or infinite mixtures [Escobar and West (1995); Wolpert and Ickstadt (1998); Scricciolo (2006); Kottas and Sansó (2007); Ji et al. (2009)]. Recently, such approaches have also been used to model spatio-temporal point processes [Kottas and Sansó (2007); Taddy and Kottas (2009); Taddy (2010); Taddy and Kottas (2012); Ding et al. (2012); Opitz et al. (2020)]. Non-parametric methods are quite flexible and can describe densities or intensity functions of point processes well; however, they are highly complex mathematically and as such, present computational challenges.

The parametric approach has also been used to model spatial point processes, particularly using a finite mixtures framework. Despite the fact that there is a plethora of papers in the literature on finite mixtures, this approach has not been very popular due to the intractable form of the likelihood, especially when the number of components is unknown. [Dempster et al. (1977) introduced the concept of data augmentation alongside the EM algorithm, which made the likelihood of finite mixture models more tractable and the estimation of the parameters of finite mixtures feasible. Furthermore, the Gibbs samplers by Tanner and Wong (1987) and Diebolt and Robert (1994) were significant developments towards building a hierarchical Bayesian framework that helped with Bayesian estimation for the parameters of a finite mixture with a fixed number of components.

The finite mixture approach is particularly useful when specific features of the intensity function are of interest, since they can provide inference for individual features of an intensity surface. In particular, Chakraborty and Gelfand (2010) considered a finite Gaussian mixture for a small and fixed number of components for the intensity function of a NHPPP. This idea was further extended by Micheas (2011) to include marked non-homogeneous Poisson point processes (MNHPPPs) and, more importantly, for the case of varying-dimension finite mixtures. Herein, we extend these methods to include a time component in the intensity function in order to model DSTPPPs.

Many spatial-temporal point processes have events that may occur at any continuous
time point on the positive real line; thus, a natural model for such processes would be a temporal point process that respects the natural ordering of time. One such temporal point process is the one-dimensional (1-d) Hawkes process. The Hawkes processes introduced by Hawkes (see Hawkes, 1971a, 1971b, 1972; Hawkes and Oakes, 1974; Li et al., 2020; Zhuang and Mateu, 2019; Reinhart, 2018) is a self-exciting point process that is analytically tractable and has been used in various applications in many fields such as seismology (Hawkes, 1973; Adamopoulos, 1976; Vere-Jones and Ozaki, 1982; Ogata, 1988), criminology (Mohler et al., 2011), biology (Balderama, 2012), neurophysiology (Chornoboy et al., 1988), finance, insurance, and economics (Dassios and Zhao, 2012; Embrechts et al., 2011; Errais et al., 2010).

In addition, the Hawkes process combines, in one model, both a cluster process representation and a simple conditional intensity representation. The Hawkes process comes closest to fulfilling, for point processes, the kind of role that the autoregressive model plays for conventional time series. However, its representation as a cluster process only allows the Hawkes process to be used in situations where the point patterns are over-dispersed relative to the Poisson model (Daley and Vere-Jones, 2003, pgs 180-183).

To model a dynamic time process we propose a MNHPPP with the intensity function defined via conditioning arguments that allow for additional interpretations. We use the term dynamic in order to indicate that some of the parameters of the models evolve over time and depend on past values, typically in a Markovian fashion. More precisely, an event of the DSTPP consists of the time (temporal) and location (spatial) components with the joint intensity function decomposed in two parts. The first part is a function representing the intensity of the temporal (time) component of the event, which is responsible for the number of the events. This is multiplied by the second part, which is a function representing the intensity of the spatial (location) process defined conditionally on the time component of the event. In particular, the temporal component is modeled by the Hawkes process and the spatial (location) process is modeled using a mixture of bivariate normal components. Secondly, we propose a flexible hierarchical Bayesian framework for estimation of the parameters of the intensity function.
This chapter is organized as follows. In Section 2.2, we review briefly the proposed Hawkes process model and illustrate the model for the spatio-temporal non-homogeneous Poisson point process (STNHPPP) intensity function. Section 2.3 presents the hierarchical Bayesian formulation and the algorithms for fitting the models. Simulation details for the proposed models are given in Section 2.4. In Section 2.5, we present an application to modeling the occurrences of earthquakes in California, USA. Concluding remarks are given in the last section.

2.2 Modeling a spatial-temporal point process via conditioning

We present two major modeling approaches; first we view the points over space and time as point patterns arising from a marked space-time NHPPP and second, we directly specify a multivariate model on the points treating the number of points as fixed. We begin with the point process approach.

2.2.1 Marked space-time non-homogeneous Poisson point process models

Point patterns arising from NHPPP models are simple, meaning that all events are different from each other (i.e., no replicates). However, due to the nature of the application that we consider, we introduce models that have the capability to handle replication (i.e., point process models known as multitype or multivariate). We briefly discuss a general NHPPP.

Consider a planar region \( \mathcal{W} \subset \mathcal{R}^2 \) (extensions to higher dimensions are straightforward), and suppose that we observe \( n \) points (events) \( \{s_i\}_{i=1}^n \). In order to model this collection of points we consider an NHPPP, which assumes that the random variables (counting variables) \( N(B), B \subseteq \mathcal{W} \), are Poisson, with mean measure \( E[N(B)] = \Lambda(B) = \int_B \lambda(s)ds \), where \( \lambda(s) \) is
known as the intensity function or surface for planar point patterns and uniquely determines the distribution of \( N \).

In addition, the counts are independent over any finite collection of disjoint regions. The special case where \( \lambda(s) = \lambda \) yields the homogeneous Poisson point process (HPPP), with rate \( \lambda \), and mean measure \( \Lambda(B) = \lambda |B| \), where \( |B| \) denotes the area of \( B \).

The joint distribution of the events and the number of events, \( \mathcal{N}(W) = n \), over the window \( W \) is given by

\[
f(s_1, \ldots, s_n, n) = \frac{e^{-\Lambda(W)}}{n!} \prod_{i=1}^{n} \lambda(s_i) = \frac{\exp \left( - \int_W \lambda(s) ds \right)}{n!} \prod_{i=1}^{n} \lambda(s_i), \tag{2.1}
\]

for \( n \geq 0 \). In order to describe the point process, one needs to model its intensity \( \lambda(s) \). However, calculation of the integral requires approximations that lead to estimation procedures requiring numerical optimization and, therefore, substantial computational burden. We alleviate this problem by considering

\[
\lambda(s) = \lambda \times \lambda_1(s), \tag{2.2}
\]

with \( \lambda_1(s) \) a proper density, so that \( \int_W \lambda(s) ds = \lambda \int_W \lambda_1(s) ds = \lambda \), and as a result (2.1) becomes

\[
f(s_1, \ldots, s_n, n) = \frac{\lambda^n e^{-\lambda}}{n!} \prod_{i=1}^{n} \lambda_1(s_i). \tag{2.3}
\]

We consider a mixture modeling approach for \( \lambda_1 \) as follows; let \( \varphi(s|\mu, \Sigma) \) denote the density of a \( \mathcal{N}_2(\mu, \Sigma) \) distribution and define the \( m \)-component mixture of bivariate normals

\[
\lambda_1(s|\theta^{(m)}) = \sum_{j=1}^{m} p_j \varphi(s|\mu_j, \Sigma_j), \tag{2.4}
\]

where \( \theta^{(m)} = (\mu, \Sigma, \mu_1, \ldots, \mu_m, \Sigma_m) \) are the parameters of the \( m \)-component mixture, \( \mu = (p_1, \ldots, p_m) \in \mathcal{S}_m = \{ \mu : p_j \geq 0 \text{ and } \sum_{j=1}^{m} p_j = 1 \} \), are the component probabilities, and
\(\varphi_j(s|\mu_j, \Sigma_j)\) is the \(j-\text{th}\) component density. Now, \(\lambda > 0\) is interpreted as the average number of points over the region \(W\), provided that \(\int_{W} \varphi_j(s|\mu_j, \Sigma_j)ds = 1\), for all \(j = 1, 2, \ldots, m\). When a component has positive probability outside \(W\) (edge effects), then the component can be truncated to integrate to one, by letting

\[
\Phi_j(\mu_j, \Sigma_j) = P(s \in W|\mu_j, \Sigma_j) = \int_{W} \varphi_j(s|\mu_j, \Sigma_j)dx,
\]

and then replacing \(\varphi_j(s|\mu_j, \Sigma_j)\) in (2.4) by \(\varphi_j(s|\mu_j, \Sigma_j)/\Phi_j(\mu_j, \Sigma_j)\). It is not hard to implement such formulations in a Bayesian setting since we only need to introduce a Metropolis-Hastings step in the MCMC sampler and the only drawback is an additional computational burden at each iteration. Since many applications have edge effects we can expand \(W\) as much as we want in order to have \(\varphi_j(s|\mu_j, \Sigma_j)\) proper over \(W\) for all \(j\), and hence, we can assume that \(\Phi_j(\mu_j, \Sigma_j) = 1, j = 1, 2, \ldots, m\).

The time or mark component is introduced in the event by simply augmenting the event space. That is, first we extend the NHPPP \(N = \{s : s \in W\}\) to the Marked NHPPP, \(N_\xi = \{(s, \xi) : s \in W, \xi \in \mathcal{M}\}\), where \(\mathcal{M}\) denotes the mark space. Then, we include the time component by further extending \(N_\xi\) to the marked space-time NHPPP, \(N_{t\xi} = \{(s, \xi, t) : s \in W, \xi \in \mathcal{M}, t \in [0, T]\}\). This formulation is straightforward, since \(N_\xi\) and \(N_{t\xi}\) are (unmarked) point process on the spaces \(W \times \mathcal{M}\) and \(W \times \mathcal{M} \times [0, T]\), respectively. The intensity of the marked space-time NHPPP, \(N_{t\xi}\), can be modeled via conditioning arguments as we see below.

We describe each one of the important components of our modeling approach below along with some of the possible models we can build in this context, starting with the model for the time component of the event of a multitype marked NHPPP.

For the temporal component of a space-time point process is modeled using a one dimensional (1-d) Hawkes process [Hawkes (1971b)]. This process can be viewed as a self-exciting non-homogeneous Poisson point process whose instantaneous event rate depends on the history of the process, denoted by \(\mathcal{H}_t\) (i.e., on all past events). Moreover, it is a point process model that adheres to the natural ordering of time.
The instantaneous rate of events at time $t$, denoted by $h(t|\mathcal{H}_t)$, is given

$$h(t|\mathcal{H}_t) = u(t) + \sum_{i: t_i < t} g(t - t_i)$$  \hspace{1cm} (2.5)$$

where $t_i$ denotes the $i$ -th event, $u(t)$ is the response function, and $g(t)$ is the memory kernel (i.e., the extra rate incurred by an event). Depending on the application, the memory kernel $g(t)$ can take a variety of forms such as hyperbolic function (i.e., the power law) or a superposition of the gamma function. However, in our case, we consider an exponentially decaying memory kernel $g(t) = \alpha \beta e^{-\beta t}$. The response function is taken to be constant in time, i.e., $u(t) = \mu$, which leads to

$$h(t|\mu, \alpha, \beta, \mathcal{H}_t) = \mu + \alpha \beta \sum_{i: t_i < t} e^{-\beta(t-t_i)},$$  \hspace{1cm} (2.6)$$

where $\mathcal{H}_t$ is the history of the process, $\mu > 0$ is the initial intensity (i.e., the intensity at time $t = 0$), and $\alpha \in (0, 1)$ is the size of the self exciting jumps relative to the constant exponential decay rate $\beta > 0$.

Now, denote by $\psi_K = \{t_1, ..., t_K\}$ a point pattern of $K$ temporal events over the time interval $\mathcal{T} = [0, T)$, and write $\psi_K \sim Hawkes(\mu, \alpha, \beta, T)$. The joint distribution of $\psi$ is given by

$$f(\psi_K|\mu, \alpha, \beta) = e^{-\int_{t=0}^{T} h(t|\mu, \alpha, \beta, \mathcal{H}_t) dt} \prod_{k=1}^{K} h(t_k|\mu, \alpha, \beta, \mathcal{H}_{t_k}),$$

or

$$f(\psi_K|\mu, \alpha, \beta) = e^{-\Lambda(\mu, \alpha, \beta, T)} \prod_{k=1}^{K} \left( \mu + \alpha \beta \sum_{i: t_i < t_k} e^{-\beta(t_k-t_i)} \right),$$  \hspace{1cm} (2.7)$$

where $\Lambda(\mu, \alpha, \beta, T) = \int_{t=0}^{T} h(t|\mu, \alpha, \beta, \mathcal{H}_t) dt$ must be computed at each iteration of the posterior sampler of $\mu, \alpha$ and $\beta$.

A STPP $\mathcal{N}$ is typically defined as a counting measure over some bounded region $(\mathcal{W} \times \mathcal{T})$, where $\mathcal{W} \subseteq \mathbb{R}^2$ and $\mathcal{T} \subseteq \mathbb{R}^+$. We can represent $\mathcal{N}$ equivalently as a collection of (distinct) random points $\{x_1 = (s_1, t_1), ..., x_n = (s_n, t_n)\}$, where $\mathcal{N}(\mathcal{W} \times \mathcal{T}) = n$ represents the total
number of points of the process \( N \) over the region \((W \times T)\) and where each event \( x_i \) contains information about the spatial location \( s_i \) and the time \( t_i \) associated with the \( i-th \) event (see Ogata, 1998; Diggle, 2006; Mohler et al., 2011; Raeisi et al., 2021).

Now, to define the STPP we consider a Poisson process with intensity function \( \lambda(s, t) \), \( s \in W \) and \( t \in [0, T] \subseteq T \). To model the STPP via conditioning, one may choose different types of conditioning in describing \( \lambda(s, t) \) depending on the nature of the problem and the questions that one seeks to address. For example, when the data consist of a collection of points (locations) observed over a specific window at given times, then it is makes sense to condition on the time component. In contrast, if we have locations that we visit over time (e.g., checking a certain spot in a forest for a specific species of bird), then conditioning on the spatial component is more appropriate. In general, we can write the joint intensity via conditioning in any of the following forms

\[
\lambda(s, t|\theta) = \lambda \times \lambda_1(t|\theta_1) \times \lambda_2(s|t, \theta_2(t)), \quad (2.8)
\]

\[
\lambda(s, t|\theta) = \lambda \times \lambda_3(s|\theta_3) \times \lambda_4(t|s, \theta_2(s)), \quad (2.9)
\]

where \( \lambda > 0 \) is some scaling parameter, \( \lambda_1, \lambda_2, \lambda_3 \) and \( \lambda_4 \) are proper densities, and

\[
\Lambda(W \times T) = \int_T \int_{t=0}^T \lambda(s, t)dsdt = \lambda \int_T \int_{t=0}^T \lambda_1(t|\theta_1) \left[ \int_W \lambda_2(s|t, \theta_2(t))ds \right] dt = \lambda,
\]

is interpreted as the average number of events occurring over the window \( W \times T \) (i.e., the average total number of events).

In this chapter we will concentrate on conditioning on the time component, considering the first type of conditioning in (2.8) we notice that for each temporal event \( t_i \) there is a single spatial event at location \( s_i \) giving rise to a point pattern of the form

\[
\varphi_n = \{(s(t_i), t_i), \ i = 1, \ldots, n\}. \quad (2.10)
\]
Therefore, the joint distribution of $\varphi_n$ and $n$ is given by

$$
 f(\varphi_n, n|\theta) = \frac{e^{-\Lambda(W \times T)}}{n!} \prod_{i=1}^{n} \lambda(s_i, t_i|\theta) = \frac{\lambda^n e^{-\lambda}}{n!} \prod_{i=1}^{n} \lambda_1(t_i|\theta_1) \lambda_2(s_i(t_i)|t_i, \theta_2(t_i)),
$$

(2.11)

where $\theta$ denotes all the parameters of the STPP model.

Defining marked NHPPP models via conditioning is straightforward, following similar steps as above. More precisely, the point pattern in this case is of the form

$$
 \varphi_n = \{(s_i, \xi_i), \ i = 1, \ldots, n, \ s_i \in \mathcal{W}, \ \xi_i \in \mathcal{M}\},
$$

(2.12)

and it can be modeled as a marked NHPPP using the intensity function $\lambda(s, \xi)$. Depending on the type of data, we condition either on the mark value or the location of the event. That is, we can model the joint intensity function using

$$
 \lambda(s, \xi|\theta) = \lambda \times \lambda_1(\xi|\theta_1) \times \lambda_2(s|\xi, \theta_2(\xi)),
$$

(2.13)

$$
 \lambda(s, \xi|\theta) = \lambda \times \lambda_3(s|\theta_3) \times \lambda_4(\xi|s, \theta_2(s)),
$$

(2.14)

where $\lambda_1$, $\lambda_2$, $\lambda_3$ and $\lambda_4$ are proper densities and $\theta$ denotes all the parameters of the STPP model.

We concentrate on the model of Equation (2.13). Let $\mathcal{M} = \{\gamma_1, \ldots, \gamma_K\}$ denote a mark space with $K$ discrete marks; i.e., consider a multi-type point process. In this case the point pattern is given by

$$
 \varphi_n = \{(s_i(\xi_i), \xi_i), \ i = 1, \ldots, n, \ s_i \in \mathcal{W}, \ \xi_i \in \mathcal{M}\},
$$

(2.15)
which leads to the joint distribution

\[ f(\varphi_n, n|\theta) = \frac{e^{-\Lambda(W \times M)}}{n!} \prod_{i=1}^{n} \lambda(s_i, \xi_i|\theta) \]

\[ = \frac{\lambda^n e^{-\lambda}}{n!} \prod_{i=1}^{n} \lambda_1(\xi_i|\theta_1) \lambda_2(s_i(\xi_i)|\xi_i, \theta_2(\xi_i)), \]

since

\[ \Lambda(W \times M) = \sum_{\xi \in M} \int_{\xi \in W} \lambda(s, \xi) ds = \sum_{\xi \in M} \lambda_1(\xi|\theta_1) \left[ \int_{\xi \in W} \lambda_2(s(\xi)|\xi, \theta_2(\xi)) ds \right] = \lambda, \]

where \( \theta \) denotes all the parameters of the marked NHPPP model.

To introduce replication with respect to the time component of the event, we consider the following conditional approach. First, note that the point patterns under consideration are of the form

\[ \varphi_n = \{(s_r(t_k), t_k), \quad r = 1, \ldots, n(t_k), \quad k = 1, \ldots, K, \quad s_i \in W, \quad t_k \in T = [0, T)\}, \]

and cannot be generated via a standard space-time NHPPP, since such processes are simple and will never produce replication. Instead, we consider a multitype marked NHPPP conditionally on the realization (a point pattern) from a Hawkes process that yields the marks treated as the time components of the events.

The construction is hierarchical. First, assume that the mark space \( M = \{t_1, \ldots, t_K\} \) consists of the times we wish to replicate so that the marked point pattern is of the form

\[ \varphi_n = \{(s_i(\xi_i), \xi_i), \quad i = 1, \ldots, n, \quad s_i \in W, \quad \xi_i \in \{t_1, \ldots, t_K\}\}, \]

where \( n = \sum_{k=1}^{K} n(t_k) \), and it can be treated as a realization from a multi-type marked NHPPP. Further, suppose that the observed mark values are a point pattern \( \psi_K = \{t_1, \ldots, t_K\} \) from a Hawkes process with joint distribution \( f(\psi_K|\mu, \alpha, \beta) \) given by (2.7). As a result, we can
write the joint distribution of the data \{\varphi_n, \psi_K\} as

\[
f(\varphi_n, n, \psi_K|\theta) = f(\varphi_n, n|\psi_K, \mu, \alpha, \beta, \theta_1(\psi_K)) f(\psi_K|\mu, \alpha, \beta),
\]

(2.20)

where \(\theta\) represents all the parameters of the model, \(\theta_1(\psi_K)\) are the parameters of the marked NHPPP conditional on \(\psi_K\), and the density of the marked point pattern \(f(\varphi_n, n|\psi_K, \mu, \alpha, \beta, \theta_1(\psi_K))\) is of the form (2.17); i.e.,

\[
f(\varphi_n, n|\psi_K, \mu, \alpha, \beta, \theta_1(\psi_K)) = \frac{\lambda^n e^{-\lambda} \prod_{k=1}^K n(t_k)}{n!} \prod_{r=1}^\lambda \lambda_2(\mathbf{s}_r(t_k)|\xi = t_k, \theta_2(t_k)) \lambda_1(\xi = t_k|\mathbf{q}(\psi_K)).
\]

(2.21)

The marks \(\xi\) are modeled according to the discrete mark distribution

\[
\lambda_1(\xi|\mathbf{q}(\psi_K)) = \sum_{k=1}^{K} q_k I(\xi = t_k),
\]

(2.22)

with \(\mathbf{q}(\psi_K) = (q_1, \ldots, q_K)\) and \(\sum_{k=1}^{K} q_k = 1\). The probability of observing the \(k\)th mark value (i.e., the time stamp \(t_k\)) is modeled via a discrete distribution that depends on the Hawkes process rates by defining

\[
q_k = \frac{1}{c} h(t_k|\mu, \alpha, \beta, \mathcal{H}_{t_k}),
\]

(2.23)

with \(c = \sum_{j=1}^{K} h(t_j|\mu, \alpha, \beta, \mathcal{H}_{t_j})\), the normalizing constant. Consequently, based on (2.20) we can write the joint distribution for the point pattern (2.19) as

\[
f(\varphi_n, n, \psi_K|\theta) = \frac{\lambda^n e^{-\lambda}}{n!} \prod_{k=1}^K n(t_k) \prod_{r=1}^\lambda \lambda_2(\mathbf{s}_r(t_k)|\xi = t_k, \theta_2(t_k)) \lambda_1(\xi = t_k|\mathbf{q}(\psi_K))
\]

\[
\times e^{-\Lambda(T)} \prod_{k=1}^K \left( \mu + \alpha \beta \sum_{i:t_i < t_k} e^{-\beta(t_k-t_i)} \right)
\]

\[
= e^{-\Lambda(T)} \frac{\lambda^n e^{-\lambda}}{n!} \prod_{k=1}^K q_k^{n(t_k)} \left( \mu + \alpha \beta \sum_{i:t_i < t_k} e^{-\beta(t_k-t_i)} \right) \prod_{r=1}^\lambda \lambda_2(\mathbf{s}_r(t_k)|\xi = t_k, \theta_2(t_k))
\]

36
where, using (2.4), we model the locations process by

\[ \lambda_2(s(t) | \xi = t, \theta_2(t)) = \sum_{j=1}^{m(t)} p_j(t) \varphi_{j,t}(s(t) | \mu_j(t), \Sigma_j(t)), \tag{2.24} \]

where \( m(t) \) the number of components at time \( t \), \( \mu_j(t) \) the \( j \)-th component mean vector, \( \Sigma_j(t) \) the \( j \)-th component covariance matrix, \( p_j(t) \) the \( j \)-th component probability and \( \varphi_{j,t}(s(t) | \mu_j(t), \Sigma_j(t)) \) denotes the bivariate normal density \( N_2(\mu_j(t), \Sigma_j(t)) \). Finally, using (2.22) and (2.23) we have

\[ f(\varphi_n, n, \psi_K | \theta) = e^{-\Lambda(T)} \frac{\lambda^n e^{-\lambda}}{n!} \prod_{k=1}^{K} q_k^{n(t_k)} \left( \mu + \alpha \beta \sum_{i,k < t_k} e^{-\beta(t_k-t_i)} \right) \]
\[ \times \prod_{k=1}^{K} \prod_{r=1}^{n(t_k)} \prod_{j=1}^{m(t_k)} p_j(t_k) \varphi_{j,t_k}(s(t_k) | \mu_j(t_k), \Sigma_j(t_k)), \tag{2.25} \]

with the product of the sum above introducing an intractability that needs to be addressed in order to facilitate Bayesian estimation.

Given the value of the time component \( t \), the spatial part of the marked NHPPP is given by the intensity function of Equation (2.24), which leads to the intractable likelihood of Equation (2.25). Following Dempster et al. (1977), any mixture model can be thought of as a missing or incomplete data model, where for each observed \( s_i(t) \) we define a vector of latent variables \( z_i(t) = [z_{i,1}(t), \ldots, z_{i,m(t)}(t)]^T \), which indicates to which component \( s_i(t) \) belongs (membership indicator). More precisely, \( z_{i,j}(t) = 1 \) if the \( i \)-th event at time \( t \) comes from the \( j \)-th component, or 0 otherwise, with \( \sum_{j=1}^{m(t)} z_{i,j}(t) = 1 \), for each \( i = 1, \ldots, n(t) \). Note that \( p(z_{i,j}(t_k) = 1 | \theta_2(t)) = p_j(t) \) and \( f(s_i(t) | z_{i,j}(t) = 1, \theta_2(t)) = \varphi_{j,t}(s_i(t) | \mu_j(t), \Sigma_j(t)) \), for \( i = 1, \ldots, n(t) \) and \( j = 1, \ldots, m(t) \). In the point process context we can think of \( z_i(t) \) as marks associated with \( s_i(t) \) so that \( z_i(t) \in M_1 = \{ z_i(t) = [z_{i,1}(t), \ldots, z_{i,m(t)}(t)]^T : z_{i,j}(t) \in \{0,1\} \text{ and } \sum_{j=1}^{m(t)} z_{i,j}(t) = 1 \} \), where \( M_1 \) is the mark space. Augmenting each \( s_i(t) \) by \( z_i(t) \) in the point pattern observed at time \( t \), we obtain the “completed intensity function” for model
for the locations, so that (2.24) reduces to

\[ \lambda_2^*(s_i(t), z_i(t)|\xi = t, \theta_2(t)) = \prod_{j=1}^{m(t)} p_j(t)^{z_{ij}(t)} \varphi_{j,t}(s_i(t)|\mu_j(t), \Sigma_j(t))^{z_{ij}(t)}, \]  

(2.26)

which alleviates the intractability issue in (2.25) (the likelihood).

Combining (2.25) and (2.26) yields the data-augmented likelihood for the marked space-time NHPPP. More precisely, all inference (Bayesian or otherwise) for a point pattern \( \{\varphi_n, \psi_K\} \) is based on the model

\[ f(\varphi_n, n, \psi_K|\theta) = e^{-\Lambda(\mu, \alpha, \beta, T)} \prod_{k=1}^{K} \int_{t_k}^{m(t_k)} \prod_{j=1}^{n(t_k)} p_j(t_k)^{z_{r,j}(t_k)} [\varphi_{j,t_k}(s_r(t_k)|\mu_j(t_k), \Sigma_j(t_k))]^{z_{r,j}(t_k)}, \]  

(2.27)

where \( \theta \) denotes all the parameters of the model, and \( z = \{z_r(t_k), r = 1, 2, \ldots, n(t_k), k = 1, 2, \ldots, K\} \) denotes all the data-augmentation vectors. Even though the \( z \) are part of the data in the augmented model for the marked NHPPP, they are typically unknown and will be treated as auxiliary variables requiring estimation. More precisely, replacing \( \lambda_2(s_r(t_k)|\xi = t_k, \theta_2(t_k)) \) in (2.21) for \( f(\varphi_n, n|\psi_K, \mu, \alpha, \beta, \theta_1(\psi_K)) \) with \( \lambda_2^*(s_i(t), z_i(t)|\xi = t, \theta_2(t)) \) of equation (2.26) yields the data-augmented model denoted by \( f(\varphi_n, n, z|\psi_K, \mu, \alpha, \beta, \theta_1(\psi_K)). \)

### 2.2.2 A multivariate space-time modeling approach

We consider the number of points to be fixed so that the problem reduces to typical multivariate modeling. Assume that we observe a collection of points \( \varphi_n \) in space and time given by (2.18) and we require a multivariate model to describe these data. Let \( n = \{n(t_1), \ldots, n(t_K)\}^T \) denote the numbers of points at each time period, \( t = \{t_1, \ldots, t_K\}^T \), the times observed, and \( L = \{s_r(t_k) : r = 1, \ldots, n_k, k = 1, \ldots, K\} \), the locations. Using
conditional arguments we can write the joint distribution of \((L, n, t)\) as

\[
f(L, n, t|\mu, \alpha, \beta, \theta) = f(L, n|\theta, \mu, \alpha, \beta, t)f(t|\mu, \alpha, \beta) = f(L|\theta_1, n, t)f(n|\mu, \alpha, \beta, t)f(t|\mu, \alpha, \beta)
\]

\[
= \left[ \prod_{k=1}^{K} \prod_{r=1}^{n_k} f(s_r(t_k)|\theta_1, t_k) \right] \left[ \prod_{k=1}^{K} p(n(t_k)|\mu, \alpha, \beta, t_k) \right] f(t|\mu, \alpha, \beta),
\]

where \(\theta\) denotes all the parameters of the model with \(f(L, n, t|\mu, \alpha, \beta, \theta)\) a proper density. Therefore, we need to model the densities \(f(s_r(t_k)|\theta_1, t_k), p(n(t_k)|\mu, \alpha, \beta, t_k)\) and \(f(t|\mu, \alpha, \beta)\).

In order to adhere to the natural ordering of time we choose a Hawkes process as the model that yields the time stamps, i.e., \(t|\mu, \alpha, \beta, T \sim Hawkes(\mu, \alpha, \beta, T)\), so that \(f(t|\mu, \alpha, \beta)\) is given by (2.7). The model for the number of locations \(n_k\) observed at time \(t_k\) is taken to be a Poisson distribution with some intensity \(\gamma(t_k)\) at time \(t_k\), that is, \(n(t_k)|\gamma(t_k), t_k \sim Poisson(\gamma(t_k))\) and we have

\[
p(n(t_k)|\gamma(t_k), t_k) = e^{-\gamma(t_k)} \frac{[\gamma(t_k)]^{n(t_k)}}{n(t_k)!}.
\]

The last model required is that for the locations. We entertain again a mixture model similar to (2.24) (or its data augmented version of equation (2.26) to remove the likelihood intractability). Specifically,

\[
f(s_r(t_k)|\theta_{1k}, t_k) = \sum_{j=1}^{m(t_k)} p_j(t_k) \varphi_{j, t_k}(s_r(t_k)|\mu_j(t_k), \Sigma_j(t_k)),
\]

where \(\theta_{1k}\) denotes all the parameters of the mixture model at time \(t_k\), which depend on time and will be evolved dynamically. Finally, using the augmented distribution for the locations and the data augmentation vectors \(z\) (see Section 2.2.1 and equation (2.26)), we can write
the data-augmented joint distribution of $\varphi_n$ and $z$ as

$$f(L, n, t, z|\mu, \alpha, \beta, \theta) = \prod_{k=1}^{K} \prod_{r=1}^{n_k} \prod_{j=1}^{m(t_k)} p_j(t_k)^{z_{r,j}(t_k)} [\varphi_{j,t_k}(s_{r}(t_k)|\mu_j(t_k), \Sigma_j(t_k))]^{z_{r,j}(t_k)}$$

$$\times \prod_{k=1}^{K} \frac{e^{-\gamma(t_k)} \gamma(t_k)^{n(t_k)}}{n(t_k)!}$$

$$\times e^{-\Lambda(\mu, \alpha, \beta, T)} \prod_{k=1}^{K} \left( \mu + \alpha \beta \sum_{i:t_i < t_k} e^{-\beta(t_k-t_i)} \right),$$

(2.28)

with $\theta = \{\theta_1(t_k), \gamma(t_k)\}_{k=1}^{K}, \theta_1(t_k) = \{m(t_k), p(t_k), \{\mu_j(t_k)\}_{j=1}^{m(t_k)}, \{\Sigma_j(t_k)\}_{j=1}^{m(t_k)}\}$ and $p(t_k) = [p_1(t_k), \ldots, p_{m(t_k)}(t_k)]^T$, and all inference is based on this data model.

### 2.3 Hierarchical Bayesian Formulations

We present the general hierarchical Bayesian formulation for both models, the marked space-time NHPPP and the space-time multivariate model. The general structure of Bayesian hierarchical models (Berliner, 1996; Wikle et al., 1998; Cressie and Wikle, 2015) can be described as follows:

**Stage 1. Data Model:** At the highest level we observe the data from some distribution [data|process, data parameters]; that is we assume that the observed data is a result of some known process and known parameters.

**Stage 2. Process Model:** The process that governs the observed data is itself dependent on some parameters; that is, it is modeled using some distribution [process|process parameters].

**Stage 3. Parameter Model:** The parameters of the data and process models are modeled according to some prior distribution [data and process parameters|hyper-parameters].

Stage 2 and 3 can be thought of as the prior stage in the absence of a process. We can include additional prior stages on the hyper-parameters of the parameter models or keep them fixed. In particular, fixed hyper-parameters were chosen in order to make the priors have the least influence on the posterior.
2.3.1 Stage 1: The Data Models

We denote the point pattern observed at time $t_k$ by $\zeta_k = \{(s_r(t_k), t_k)\}_{r=1}^{n(t_k)}$, for $k = 1, \ldots, K$, and the spatial-temporal pattern by $\varphi_n = \{\zeta_k\}_{k=1}^{K}$, with $n = \sum_{k=1}^{K} n(t_k)$. The data model for the marked space-time NHPPP model is given by (2.27), and for the multivariate space-time model by (2.28).

Note that both modeling approaches require non-dynamic modeling of the Hawkes parameters $\mu, \alpha, \beta$, and dynamic modeling of the evolving mixture parameters $m(t_k), p_j(t_k), \mu_j(t_k)$ and $\Sigma_j(t_k)$, for $j = 1, 2, \ldots, m(t_k)$ and $k = 1, \ldots, K$. The only differences are that the marked space-time NHPPP has an additional parameter $\lambda$ that requires modeling, whereas, the multivariate space-time model requires dynamic modeling of $\gamma(t_k)$.

2.3.2 Stage 2: Process stage

The process stages for both the multi-type model and the multivariate model are similar thus, what follows applies to both models. Note that even though the data augmentation vectors $z_r(t_k)$ depend on time, they are not dynamically evolving, so that they can be estimated via a standard data-augmentation Gibbs sampler (Diebolt and Robert [1994]).

The only parameters that are dynamically evolving at stage 1 (Data stage for both models) are parameters $\theta = \{\theta_1(t_k), \gamma(t_k)\}_{k=1}^{K}$, with the mixture parameters given by

$$\theta_1(t_k) = \left\{m(t_k), p(t_k), \{\mu_j(t_k)\}_{j=1}^{m(t_k)}, \{\Sigma_j(t_k)\}_{j=1}^{m(t_k)}\right\},$$

and $p(t_k) = [p_1(t_k), \ldots, p_{m(t_k)}(t_k)]^T$, with $\gamma(t_k)$ appearing only in the marked space-time NHPPP model. Since all the parameters, including the number of components, depend on time $t$ we model each one of them using dynamic evolution models. Specifically, each of
these parameters is modeled dynamically across time using
\[
\log \gamma(t_k) = \log(\gamma(t_{k-1})) + \epsilon_\gamma(t_k), \quad \epsilon_\gamma(t_k) \sim N(0, \sigma_\gamma^2),
\]
(2.30)
\[
m(t_k)|\tau(t_k) \sim \text{truncated Poisson}(\tau(t_k)) \text{ at 1},
\]
(2.31)
\[
\pi(z_r(t_k)|m(t_k)) \propto 1, \text{ for all } r = 1, 2, \ldots, n(t_k),
\]
(2.32)
\[
p(t_k)|d(t_k), m(t_k) \sim \text{Dirichlet}(d(t_k)),
\]
(2.33)
\[
\mu_j(t_k)|\psi(t_k), A \sim N(\psi(t_k), A), \quad \text{and}
\]
\[
\Sigma_j^{-1}(t_k)|a_1, B(t_k) \sim \text{Wishart}_2(2a_1, (B(t_k))^{-1}),
\]
(2.35)
where \(d(t_k) = [d_1(t_k), \ldots, d_m(t_k)]^T\), with \(\log(\gamma(t_1))|\tilde{\mu}_\gamma, \tilde{\sigma}_\gamma^2 \sim N(\tilde{\mu}_\gamma, \tilde{\sigma}_\gamma^2)\), for some fixed \(\tilde{\mu}_\gamma, \tilde{\sigma}_\gamma^2\) at the initial time period \(t_1\). To complete the hierarchical formulation, we require evolution models for \(\tau(t_k), d(t_k), \psi(t_k), \Psi(t_k)\) and \(B(t_k)\). Note that by construction, conditioned on \(t_k\), the dimension of the parameter space for these parameters remains constant since they do not depend on the specific component \(j = 1, 2, \ldots, m(t_k)\), only the time period \(t_k\). Therefore, a fixed dimension Gibbs sampler is required to sample from the posterior of these parameters using Metropolis-within-Gibbs steps as needed. If we let \(j\) affect these parameters, then we would require a variable dimension sampler, like Reversible Jump MCMC (Green 1995) or Birth-Death MCMC (Stephens 2000).

**Dynamic evolution of the process parameter \(\tau(t_k)\)**

Since \(\tau(t_k)\) are positive we consider an AR(1) process for \(\log(\tau(t_k))\) following Dey and Micheas (2014); i.e., \(\log(\tau(t_k)) = \varphi_\tau \log(\tau(t_{k-1})) + \epsilon_1(t_k), \quad \epsilon_1(t_k) \sim N(0, \sigma_\tau^2)\). Therefore, we have that \(\log(\tau(t_k))|\tau(t_{k-1}), \varphi_\tau, \sigma_\tau^2 \sim N(\varphi_\tau \log(\tau(t_{k-1})), \sigma_\tau^2)\) and we set \(\log(\tau(t_1))|\tilde{\mu}_\tau, \tilde{\sigma}_\tau^2 \sim N(\tilde{\mu}_\tau, \tilde{\sigma}_\tau^2)\), for some fixed \(\tilde{\mu}_\tau, \tilde{\sigma}_\tau^2\) at the initial time period \(t_1\). We simplify the problem by considering a random walk; i.e., we set \(\varphi_\tau = 1\).
Dynamic evolution of the process parameter vector $d(t_k)$

Now, notice that $d(t_k) = [d_1(t_1), \ldots, d_{m(t_k)}(t_k)]^T$, $d_j(t_k) > 0$, is a vector of length $m(t_k)$ and thus, its dimension is changing across time. To model each component of $d(t_k)$ we assume that $\log(d_j(t_k)) \sim N(\mu_d(t_k), \sigma^2_d)$, for $j = 1, \ldots, m(t_k)$, $k = 1, \ldots, K$, so that $d_j(t_k)$ have a common mean parameter, $\mu_d(t_k)$, that evolves with time and a non-evolving variance parameter, $\sigma^2_d$. We now consider an AR(1) process for $\mu_d(t_k)$; that is, $\mu_d(t_k) = \varphi_d \mu_d(t_{k-1}) + \epsilon_2(t_k)$, $\epsilon_2(t_k) \sim N(0, \sigma^2_d)$. As a result $\mu_d(t_k)|\mu_d(t_{k-1}), \varphi_d, \sigma^2_d \sim N(\varphi_d \mu_d(t_{k-1}), \sigma^2_d)$, where for the initial time, we take $\mu_d(t_1)|\mu_d, \sigma^2_d \sim N(\tilde{\mu}_d, \tilde{\sigma}^2_d)$, for some fixed $\tilde{\mu}_d$, and $\tilde{\sigma}^2_d$. Again, consider a random walk and set $\varphi_d = 1$. Note that even though $m(t_k)$ may change over time, the approach we take yields a fixed-dimension parameter space so that fixed-dimension samplers can be used to sample from the posterior distribution.

Dynamic evolution of the process parameter vector $\psi(t_k)$

We now consider a VAR(1) model for $\psi(t_k)$; that is, $\psi(t_k) = \Pi \psi(t_{k-1}) + \eta(t_k)$, $\eta(t_k) \sim N_2(0, \Delta(t_k))$. Thus, we have $\psi(t_k)|\Pi, \psi(t_{k-1}), \Delta(t_k) \sim N_2(\Pi \psi(t_{k-1}), \Delta(t_k))$ and set $\psi(t_1)|\bar{\mu}_\psi, \bar{\Sigma}_\psi \sim N_2(\bar{\mu}_\psi, \bar{\Sigma}_\psi)$, for some fixed $\bar{\mu}_\psi$ and $\bar{\Sigma}_\psi$, with $\psi(t_1) = [\psi_1(t_k), \psi_2(t_k)]^T$ and $\Pi = \text{diag}(\pi_1, \pi_2)$. Specifically, we consider a random walk by taking $\Pi = I_2$, the identity matrix.

Now in order to build an evolution model for $\Delta(t_k)$, we consider the model proposed by Dey and Micheas (2014); that is, we decompose $\Delta(t_k)$ using the special LDU decomposition method originally proposed by Dunson et al. (2003) and further developed by Chen et al. (2011) as follows

$$\Delta(t_k) = K(t_k) \Gamma(t_k) \Gamma(t_k)^T K(t_k), \quad (2.36)$$

where $K(t_k) = \text{diag}(\kappa_1(t_k), \kappa_2(t_k))$ and $\Gamma(t_k) = \begin{pmatrix} 1 & 0 \\ \zeta(t_k) & 1 \end{pmatrix}$. Instead of evolving the variance-covariance matrix $\Delta(t_k)$, we evolve $\kappa_1(t_k), \kappa_2(t_k)$ and $\zeta(t_k)$, individually. Since $\Delta(t_k)$ is positive definite, the values of $\kappa_1(t_k)$ and $\kappa_2(t_k)$ must be positive. Therefore, we consider AR(1) processes for $\log(\kappa_1(t_k))$, $\log(\kappa_2(t_k))$ and $\zeta(t_k)$ as above. More precisely, we
take

\[
\log(\kappa_1(t_k)) = \varphi_{\kappa_1} \log(\kappa_1(t_{k-1})) + \epsilon_4(t_k), \ \epsilon_4(t_k) \sim N(0, \sigma_{\kappa_1}^2), \\
\log(\kappa_2(t_k)) = \varphi_{\kappa_2} \log(\kappa_2(t_{k-1})) + \epsilon_5(t_k), \ \epsilon_5(t_k) \sim N(0, \sigma_{\kappa_2}^2), \\
\zeta(t_k) = \varphi_\zeta \zeta(t_{k-1}) + \epsilon_3(t_k), \ \epsilon_3(t_k) \sim N(0, \sigma_\zeta^2),
\]

and thus, we have \( \log(\kappa_1(t_k))|\varphi_{\kappa_1}, \kappa_1(t_{k-1}), \sigma_{\kappa_1}^2 \sim N(\varphi_{\kappa_1} \log(\kappa_1(t_{k-1})), \sigma_{\kappa_1}^2) \), with initial state \( \log(\kappa_1(t_1))|\mu_{\kappa_1}, \sigma_{\kappa_1}^2 \sim N(\mu_{\kappa_1}, \sigma_{\kappa_1}^2) \), \( \log(\kappa_2(t_k))|\varphi_{\kappa_2}, \kappa_2(t_{k-1}), \sigma_{\kappa_2}^2 \sim N(\varphi_{\kappa_2} \log(\kappa_2(t_{k-1})), \sigma_{\kappa_2}^2) \), with initial state \( \log(\kappa_2(t_1))|\mu_{\kappa_2}, \sigma_{\kappa_2}^2 \sim N(\mu_{\kappa_2}, \sigma_{\kappa_2}^2) \), and \( \zeta(t_k)|\varphi_\zeta, \zeta(t_{k-1}), \sigma_\zeta^2 \sim N(\varphi_\zeta \zeta(t_{k-1}), \sigma_\zeta^2) \), with initial state \( \zeta(t_1)|\mu_\zeta, \sigma_\zeta^2 \sim N(\mu_\zeta, \sigma_\zeta^2) \). We will use a random walk by setting \( \varphi_{\kappa_1} = 1, \varphi_{\kappa_2} = 1 \) and \( \varphi_\zeta = 1 \).

**Dynamic evolution of the process parameter matrix** \( B(t_k) \)

To model \( B(t_k) \) we appeal once again to the LDU decomposition method, so that

\[
B(t_k) = P(t_k)F(t_k)F(t_k)^T P(t_k),
\]

(2.37)

where \( P(t_k) = \text{diag}(\rho_1(t_k), \rho_2(t_k)) \) and \( F(t_k) = \begin{pmatrix} 1 & 0 \\ \varrho(t_k) & 1 \end{pmatrix} \). As before, instead of evolving the variance-covariance matrix \( B(t_k) \), we evolve \( \rho_1(t_k), \rho_2(t_k) \) and \( \varrho(t_k) \) individually. Since \( B(t_k) \) is positive definite, the values of \( \rho_1(t_k) \) and \( \rho_2(t_k) \) must be positive. Therefore, we consider AR(1) processes for \( \log(\rho_1(t_k)) \), \( \log(\rho_2(t_k)) \) and \( \varrho(t_k) \), that is,

\[
\log(\rho_1(t_k)) = \varphi_{\rho_1} \log(\rho_1(t_{k-1})) + \epsilon_{10}(t_k), \ \epsilon_{10}(t_k) \sim N(0, \sigma_{\rho_1}^2), \\
\log(\rho_2(t_k)) = \varphi_{\rho_2} \log(\rho_2(t_{k-1})) + \epsilon_{11}(t_k), \ \epsilon_{11}(t_k) \sim N(0, \sigma_{\rho_2}^2), \\
\varrho(t_k) = \varphi_\varrho \varrho(t_{k-1}) + \epsilon_9(t_k), \ \epsilon_9(t_k) \sim N(0, \sigma_\varrho^2),
\]

so that \( \log(\rho_1(t_k))|\varphi_{\rho_1}, \rho_1(t_{k-1}), \sigma_{\rho_1}^2 \sim N(\varphi_{\rho_1} \log(\rho_1(t_{k-1})), \sigma_{\rho_1}^2) \), with initial state modeled by \( \log(\rho_1(t_1))|\tilde{\mu}_{\rho_1}, \sigma_{\rho_1}^2 \sim N(\tilde{\mu}_{\rho_1}, \sigma_{\rho_1}^2) \), \( \log(\rho_2(t_k))|\varphi_{\rho_2}, \rho_2(t_{k-1}), \sigma_{\rho_2}^2 \sim N(\varphi_{\rho_2} \log(\rho_2(t_{k-1})), \sigma_{\rho_2}^2) \),
with initial state modeled by \( \log(p_2(t_1)) \mid \bar{\mu}_{p_2}, \bar{\sigma}_{p_2}^2 \sim N(\bar{\mu}_{p_2}, \bar{\sigma}_{p_2}^2) \), and \( \varphi(t_k) \mid \varphi, \varrho(t_{k-1}), \sigma^2_\varrho \sim N(\varphi \varrho(t_{k-1}), \sigma^2_\varrho) \), with initial state \( \varrho(t_1) \mid \bar{\mu}_\varrho, \bar{\sigma}_\varrho^2 \sim N(\bar{\mu}_\varrho, \bar{\sigma}_\varrho^2) \). We use a random walk by setting \( \varphi_{\rho_1} = 1, \varphi_{\rho_2} = 1 \) and \( \varrho = 1 \).

### 2.3.3 Stage 3: Priors for the model and process stage parameters

The Hawkes process depends on the parameters \( \mu, \alpha, \) and \( \beta \), which are given prior distributions \( \mu \sim \text{Exp}(b_\mu), \alpha \sim \text{Exp}(b_\alpha) \) and \( \beta \sim \text{Exp}(b_\beta) \). In addition, we consider \( \lambda \sim G(a_\lambda, b_\lambda), \sigma_{\gamma}^2 \sim IG(a_\gamma, b_\gamma), \) and \( \sigma_{\tau}^2 \sim IG(a_\tau, b_\tau) \), for positive some real numbers \( a_\gamma, b_\gamma, a_\lambda, b_\lambda, a_\tau \) and \( b_\tau \), fixed so as to have no influence on the analysis (flat priors). Let \( \sigma_{\sigma}^2 \sim IG(a_d, b_d) \) and \( \sigma_{\rho}^2 \sim IG(a_d^*, b_d^*) \), for some fixed positive real numbers \( a_d, b_d, a_d^*, b_d^* \). Similarly, we let \( \sigma_{\kappa_1}^2 \sim IG(a_{\kappa_1}, b_{\kappa_1}), \sigma_{\kappa_2}^2 \sim IG(a_{\kappa_2}, b_{\kappa_2}), \) and \( \sigma_{\zeta}^2 \sim IG(a_{\zeta}, b_{\zeta}) \), some fixed positive real numbers \( a_{\kappa_1}, b_{\kappa_1}, a_{\kappa_2}, b_{\kappa_2}, a_{\zeta} \) and \( b_{\zeta} \). The prior distribution on the matrix \( \mathbf{A} \), the covariance matrix of the component means, is modeled by \( \mathbf{A}^{-1} \mid \mathbf{Q} \sim \text{Wishart}_{2}(2a_2, (2\mathbf{Q})^{-1}) \), for some fixed matrix \( \mathbf{Q} \). Finally, we let \( \sigma_{\varrho}^2 \sim IG(a_\varrho, b_\varrho), \sigma_{\rho_1}^2 \sim IG(a_{\rho_1}, b_{\rho_1}), \) and \( \sigma_{\rho_2}^2 \sim IG(a_{\rho_2}, b_{\rho_2}) \), for positive some real numbers \( a_\varrho, b_\varrho, a_{\rho_1}, b_{\rho_1}, a_{\rho_2} \) and \( b_{\rho_2} \).

### 2.3.4 Posterior distributions

The full-conditional distributions for both modeling approaches are presented next. Note that both models share the non-dynamic parameters of the Hawkes process parameters \( \mu, \alpha, \) and \( \beta \) and parameters \( \lambda, \mathbf{A} \), the dynamically evolving parameters \( \gamma(t_k), \tau(t_k), m(t_k), d(t_k), \) \( \mathbf{p}(t_k), \mu_j(t_k), \Sigma_j(t_k), \psi(t_k), \kappa_1(t_k), \kappa_2(t_k), \zeta(t_k), \rho_1(t_k), \rho_2(t_k), \varrho(t_k) \) for \( k = 1, \ldots, K \), and the hyper-parameters \( \sigma_{\gamma}^2, \sigma_{\tau}^2, \sigma_{d}^2, \sigma_{\kappa_1}^2, \sigma_{\kappa_2}^2, \sigma_{\zeta}^2, \sigma_{\rho_1}^2, \sigma_{\rho_2}^2, \sigma_{\varrho}^2 \).
Full posterior distribution for the marked space-time NHPPP model

The full posterior distribution of all the parameters $\theta = \{\lambda, \mu, \alpha, \beta, \{\theta_1(t_k)\}_{k=1}^K\}$ and the data augmentation data $z$ for the first model is given by

$$\pi(\theta, z|\varphi_n, n, \psi)|\propto f(\varphi_n, n, \psi|z)\pi(\theta)\pi(z)$$

$$\propto e^{-\Lambda(\mu, \alpha, \beta, \psi)}\frac{\lambda^{n}\exp\left(-\frac{n \lambda}{\mu + \alpha \beta \sum_{t_i < t_k} e^{-\beta(t_k-t_i)}}\right)}{n!}$$

$$\times \prod_{k=1}^K n(t_k) \prod_{k=1}^K \prod_{j=1}^{m(t_k)} p_j(t_k)^{z_{r,j}(t_k)} \left[\varphi_j(t_k|s_r(t_k)|\mu_j(t_k), \Sigma_j(t_k))\right]^{z_{r,j}(t_k)}$$

$$\times \pi(\lambda|\alpha, \beta)\pi(\mu|\beta, \sigma)\pi(\alpha|\alpha)\pi(\beta|\beta)$$

$$\times \prod_{k=1}^K \left[\pi(m(t_k)|\tau(t_k))\pi(p(t_k)|d(t_k), m(t_k)) \frac{m(t_k)}{K} \pi(\mu_j(t_k)|\psi(t_k), A)\pi(\Sigma_j^{-1}(t_k)|a_1, B(t_k))\right]$$

$$\times \pi(A|Q) \left[\prod_{k=1}^K \prod_{j=1}^{m(t_k)} \pi(d_j(t_k)|\mu_d(t_k), \sigma_d^2) \pi(\tau(t_1)|\mu_\tau, \sigma_\tau^2) \left[\prod_{k=2}^K \pi(\tau(t_k)|\tau(t_{k-1}), \sigma_\tau^2)\right]\right]$$

$$\times \pi(\psi(t_1)|\mu_\psi, \Sigma_\psi) \left[\prod_{k=2}^K \pi(\psi(t_k)|\psi(t_{k-1}), \Delta(t_k))\right] \prod_{k=1}^K \pi(B(t_k)|\rho_1(t_k), \rho_2(t_k), \theta(t_k))$$

$$\times \pi(\kappa_1(t_1)|\mu_{\kappa_1}, \sigma_{\kappa_1}^2)\pi(\kappa_2(t_1)|\mu_{\kappa_2}, \sigma_{\kappa_2}^2)\pi(\zeta(t_1)|\mu_\zeta, \sigma_\zeta^2)$$

$$\times \prod_{k=2}^K \pi(\kappa_1(t_k)|\kappa_1(t_{k-1}), \sigma_{\kappa_1}^2)\pi(\kappa_2(t_k)|\kappa_2(t_{k-1}), \sigma_{\kappa_2}^2)\pi(\zeta(t_k)|\zeta(t_{k-1}), \sigma_\zeta^2)$$

$$\times \pi(\rho_1(t_1)|\mu_{\rho_1}, \sigma_{\rho_1}^2)\pi(\rho_2(t_1)|\mu_{\rho_2}, \sigma_{\rho_2}^2)\pi(\varphi(t_1)|\mu_\varphi, \sigma_\varphi^2)$$

$$\times \prod_{k=2}^K \pi(\rho_1(t_k)|\rho_1(t_{k-1}), \sigma_{\rho_1}^2)\pi(\rho_2(t_k)|\rho_2(t_{k-1}), \sigma_{\rho_2}^2)\pi(\varphi(t_k)|\varphi(t_{k-1}), \sigma_\varphi^2)$$

$$\times \pi(\mu_d(t_1)|\mu_d, \sigma_d^2) \prod_{k=2}^K \pi(\mu_d(t_k)|\mu_d(t_{k-1}), \sigma_d^2)$$

$$\times \pi(\sigma_d^2|\alpha, \beta)\pi(\sigma_d^2|\alpha_d, \beta_d)\pi(\sigma_d^2|a_d, b_d)\pi(\sigma_d^2|a_{\kappa_1}, b_{\kappa_1})\pi(\sigma_d^2|a_{\kappa_2}, b_{\kappa_2})$$

$$\times \pi(\sigma_\varphi^2|\alpha, \beta)\pi(\sigma_\varphi^2|\alpha_{\varphi}, \beta_{\varphi})\pi(\sigma_\varphi^2|a_{\varphi}, b_{\varphi})$$
The full-conditional distribution of the Hawkes process parameters \( \mu, \alpha, \beta \) can be expressed as

\[
\pi(\mu, \alpha, \beta | \cdot) \propto e^{-\Lambda(\mu, \alpha, \beta, T)} e^{-\frac{\mu}{\nu_{\mu}} - \frac{\alpha}{\nu_{\alpha}} - \frac{\beta}{\nu_{\beta}}} \prod_{k=1}^{K} q_{k}^{n_{(tk)}} \left( \mu + \alpha \beta \sum_{i:t_{i}<t_{k}} e^{-\beta(t_{k}-t_{i})} \right),
\]

(2.42)

where \( \nu_{\mu}, \nu_{\alpha} \) and \( \nu_{\beta} \) are fixed positive real numbers. The full-conditional distribution above has a complicated form that cannot be sampled directly. Instead, we use a random walk Metropolis-within-Gibbs sampler with proposals being truncated normal distributions in \((0, \infty)\) for \( \mu \) and \( \beta \), \((0, 1)\) for \( \alpha \), with mean as the current values of the chain, where \( \sigma_{\mu} \) and \( \sigma_{\alpha} \) and \( \sigma_{\beta} \), some fixed values for the standard deviations. The rest of the full-conditionals are given in Appendix 2.A.

Full posterior distribution for the multivariate space-time model

The full posterior distribution of all the parameters \( \theta = \{ \mu, \alpha, \beta, \{ \theta_{1}(t_{k}), \gamma(t_{k}) \}_{k=1}^{K} \} \) and the data augmentation data \( z \) for the second model is given by

\[
\pi(\theta, z | L, n, t) \propto f(L, n, t | \mu, \alpha, \beta, \theta) \pi(\theta) \pi(z) \]

\[
\propto e^{-\Lambda(\mu, \alpha, \beta, T)} \prod_{k=1}^{K} \left( \mu + \alpha \beta \sum_{i:t_{i}<t_{k}} e^{-\beta(t_{k}-t_{i})} \right) \frac{e^{-\gamma(t_{k})} [\gamma(t_{k})]^{n_{(tk)}}}{n(t_{k})!} \times \prod_{k=1}^{K} \prod_{r=1}^{n_{(tk)}} \prod_{j=1}^{m_{(tk)}} p_{j}(t_{k})^{z_{r,j}(t_{k})} \left[ \varphi_{j,t_{k}}(s_{r}(t_{k}) | \mu_{j}(t_{k}), \Sigma_{j}(t_{k})) \right]^{z_{r,j}(t_{k})} \times \pi(\theta^{*}) \pi(\gamma(t_{1}) | \tilde{\mu}_{\gamma}, \tilde{\sigma}_{\gamma}^{2}) \prod_{k=2}^{K} \left[ \pi(\gamma(t_{k}) | \gamma(t_{k-1}), \sigma_{\gamma}^{2}) \right],
\]

where \( \pi(\theta^{*}) \) denotes the priors of the common parameters with the marked space-time NH-PPP model, given by the product of (2.39), (2.40) and (2.41).

The full-conditional distribution of the Hawkes process parameters \( \mu, \alpha, \beta \) is given by

\[
\pi(\mu, \alpha, \beta | \cdot) \propto e^{-\Lambda(\mu, \alpha, \beta, T)} e^{-\frac{\mu}{\nu_{\mu}} - \frac{\alpha}{\nu_{\alpha}} - \frac{\beta}{\nu_{\beta}}} \prod_{k=1}^{K} q_{k}^{n_{(tk)}} \left( \mu + \alpha \beta \sum_{i:t_{i}<t_{k}} e^{-\beta(t_{k}-t_{i})} \right),
\]

(2.44)
with the remainder of the setup identical to that of the full conditional in (2.42).

2.3.5 Simulating patterns from the two space-time models

We present the general algorithm that is used to simulate space-time data from the two proposed models.

Simulation of the marked space-time NHPPP

The general steps required to simulate point patterns from the proposed marked space-time NHPPP are given below.

1. Using their prior distributions, simulate all non-dynamic parameters and hyper-parameters of the model; that is, $\lambda, \mu, \alpha,$ and $\beta$.

2. For a given time interval $\mathcal{T} = [0, T)$, simulate a realization $\psi_K = \{t_1, \ldots, t_K\} \sim \text{Hawkes}(\mu, \alpha, \beta, T)$ from the Hawkes process using Algorithm 2 of Ogata (1988). The simulated temporal values $\psi_K$ represent the occurrence times of events over the temporal window $[0, T)$ and constitute the “history” of the Hawkes process, denoted by $\mathcal{H}_t$.

3. Compute $h(t_k|\mu, \alpha, \beta, \mathcal{H}_{t_k})$, for each $t_k$, $k = 1, \ldots, K$, the normalizing constant $c = \sum_{j=1}^{K} h(t_j|\mu, \alpha, \beta, \mathcal{H}_{t_j})$ and the probabilities of the mark distribution $\lambda_1(\xi|\mathbf{q}(\psi_K))$ based on equation (2.23). In what follows, all operations are performed given the value of $\psi_K$.

4. Simulate all the parameters at the process stage using their evolution models, e.g., for $\tau(t_k)$ we have $\log(\tau(t_k)) = \log(\tau(t_{k-1})) + \epsilon_1(t_k)$, $\epsilon_1(t_k) \sim N(0, \sigma_\tau^2)$, for some fixed $\sigma_\tau^2$ and initial state $\tau(t_1)$.

5. Using the evolution models described in Section 2.3.2 obtain the
(a) number of components $m(t_k)$ for each time of event $t_k \in \mathcal{H}_t$, using $m(t_k) \sim \text{truncated Poisson}(\tau(t_k))$, for $k = 1, \ldots, K$, with the truncation being at 1 so that we always have at least one mixture component at each time period $t_k$,

(b) component probabilities $\{p_j(t_k)\}_{j=1}^{m(t_k)}$, for $k = 1, \ldots, K$,

(c) component means $\{\mu_j(t_k)\}_{j=1}^{m(t_k)}$, for $k = 1, \ldots, K$, and the

(d) component covariances $\{\Sigma_j(t_k)\}_{j=1}^{m(t_k)}$, for $k = 1, \ldots, K$.

6. Simulate the multitype marked NHPPP;

(a) simulate the total number of points $n$ according to a $\text{Poisson}(\lambda)$,

(b) generate $n$ mark values according to the mark distribution $\lambda_1(\xi|\mathbf{q}(\psi_K))$,

(c) count the number of times each mark has occurred, i.e., obtain the $n(t_k)$, $k = 1, \ldots, K$,

(d) for each $k = 1, \ldots, K$, simulate $\{s_r(t_k), \ r = 1, \ldots, n(t_k)\}$, $n(t_k)$ locations from the mixture model with $m(t_k)$ the number of components at time $t_k$, $\mu_j(t_k)$, the $j$-th component mean vector, $\Sigma_j(t_k)$, the $j$-th component covariance matrix, and $p_j(t_k)$, the $j$-th component probability, $j = 1, 2, \ldots, m(t_k)$. Note that the Poisson process of locations is essentially an iid cluster process.

7. The space-time point pattern is given by $\varphi_n = \{(s_r(t_k), t_k), \ r = 1, \ldots, n(t_k), k = 1, \ldots, K, s_i \in \mathcal{W}, t_k \in \mathcal{T} = [0, T]\}$.

Simulation of the multivariate space-time model

We discuss the steps required in order to simulate space-time points from the multivariate model of Section 2.2.2

1. Use steps 1 and 2 as described in the previous section and simulate a realization $\psi_K = \{t_1, \ldots, t_K\} \sim \text{Hawkes}(\mu, \alpha, \beta, T)$ from the the Hawkes process.
2. Perform steps 4 and 5, as described in the previous section, in order to simulate the
dynamically evolving parameters and hyper-parameters of the model.

3. Simulate the number of locations \(n(t_k)\) observed at time \(t_k\) via \(n(t_k)|\gamma(t_k), t_k \sim \text{Poisson}(\gamma(t_k))\) for \(k = 1, \ldots, K\).

4. Perform step 6, d), as described in the previous section, in order to simulate the
locations of the events \(\{s_r(t_k), r = 1, \ldots, n(t_k)\}\), for each \(k = 1, \ldots, K\).

5. The space-time point pattern is given by \(\varphi_n = \{(s_r(t_k), t_k), r = 1, \ldots, n(t_k), k = 1, \ldots, K, s_i \in W, t_k \in T = [0, T)\}\).

2.4 Simulation example

First we generate the space-time data based on the two algorithms presented in Section
2.3.5 with the model parameters being known (true values) and then obtain their estimates
using the proposed BHM methodology. To assess the performance of the models, we then
compare the estimates with the true values of the parameters used to simulate the data.

2.4.1 Simulation study for the marked space-time NHPPP model

The simulated data was obtained using the procedure outlined in Section 2.3.5 as follows:
The choice of the parameters is arbitrary and it worked well for other choices.

1. First we set the average number of events to be \(\lambda = 10000\) and then to obtain the
history \(H_T\) we simulate the Hawkes process with \(\mu = 0.5, \alpha = 0.3, \beta = 0.3\) and
\(T = 50\); i.e., the history over the time window \([0, 50)\).

2. Perform step 2, as described in Section 2.3.5, in order to simulate a realization \(\psi_K =
\{t_1, \ldots, t_K\} \sim \text{Hawkes}(\mu, \alpha, \beta, T)\).

3. Perform step 3, as described in the Section 2.3.5, in order and the probabilities of the
mark distribution \(\lambda_1(\xi|q(\psi_K))\).
4. Simulate the number of components \( m(t_k) \) at each time \( t_k \in H_T \) using the evolution parameters \( \varphi_\tau = 1, \sigma_\tau = 0.1, \tilde{\mu}_\tau = 0.5 \) and \( \tilde{\sigma}_\tau = 0.5 \).

5. (a) Simulate the component probabilities \( p(t_k) \) at each time \( t_k \in H_T \) using the evolution parameters \( \varphi_d = 1, \sigma_d = 0.2, \tilde{\mu}_d = 0.2 \) and \( \tilde{\sigma}_d = 0.5 \).

(b) Simulate the component means \( \{\mu_j(t_k)\}_{j=1}^{m(t_k)} \) at each time \( t_k \in H_T \) using the evolution parameters \( A = 5 \times \text{diag}(2), \Pi = \text{diag}(2), \varphi_{\kappa_1} = 1, \sigma_{\kappa_1} = 0.2, \tilde{\mu}_{\kappa_1} = 0.5, \tilde{\sigma}_{\kappa_1} = 0.5, \varphi_{\kappa_2} = 1, \sigma_{\kappa_2} = 0.2, \tilde{\mu}_{\kappa_2} = 0.5, \tilde{\sigma}_{\kappa_2} = 0.5, \varphi_{\zeta} = 1, \sigma_{\zeta} = 0.2, \tilde{\mu}_{\zeta} = 0.1, \) and \( \tilde{\sigma}_{\zeta} = 0.5 \).

(c) Simulate the component covariances \( \{\Sigma_j(t_k)\}_{j=1}^{m(t_k)} \) at each time \( t_k \in H_T \), using the evolution parameters \( \varphi_{\rho_1} = 1, \sigma_{\rho_1} = 0.2, \tilde{\mu}_{\rho_1} = 0.5, \tilde{\sigma}_{\rho_1} = 0.2, \varphi_{\rho_2} = 1, \sigma_{\rho_2} = 0.2, \tilde{\mu}_{\rho_2} = 0.5, \tilde{\sigma}_{\rho_2} = 0.2 \) and \( \varphi_{\varrho} = 1, \sigma_{\varrho} = 0.2, \tilde{\mu}_{\varrho} = 0.1, \tilde{\sigma}_{\varrho} = 0.2 \).

6. The space-time point pattern is given by \( \varphi_n = \{(s_r(t_k), t_k), r = 1, \ldots, n(t_k), k = 1, \ldots, K, s_i \in \mathcal{W}, t_k \in T = [0, T)\} \).

After simulating the data, we implemented the MCMC and obtained the following results after running 10,000 iterations with a burn-in period of 2,000 iterations. We obtain the Poisson intensity surfaces based on the posterior means of the parameters at all times \( t_k \), \( k = 1, \ldots, 27 \), however, we only provide details for times \( t_1, t_{13}, t_{27} \) due the large number of tables and output. These times we chosen to represent at the beginning, midway and last time.

### 2.4.2 Recovering the temporal process

The temporal process part of the DSTPP is governed by the 1-d Hawkes process with intensity function given in Equation (2.6) of Section 2.2.1. In Table 2.1, we show summary results for the Hawkes process parameters, in particular, the true values of the parameters \( \mu, \alpha \) and \( \beta \), their posterior means and 95% credible sets. Figure 2.1 presents plots of the true intensity of the Hawkes process along with the estimated intensity based on the posterior means of the parameters. Note that all of the true parameter values are within
their corresponding 95% credible sets and the posterior means for the parameters are close to the true values. Figure 2.2 shows the trace plots, histograms and the associated autocorrelation function (ACF) plots for the Hawkes process parameters $\mu$, $\alpha$ and $\beta$.

2.4.3 Recovering the spatial (location) process

The spatial (location) process part of the DSTPP, conditional on the temporal component, is governed by the NHPPP with intensity function given in Equation (2.24). Figure 2.3 shows the 2-d plots of both the true and estimated Poisson intensity surfaces, along with the generated point patterns at times $t_1 = 2.6522$, $t_{13} = 25.3369$ and $t_{27} = 47.6197$. The estimated surface is based on the posterior means of the parameters, and will be hereafter referred to as the PME (posterior means estimated) intensity surface (comparison of the PME surfaces with the posterior mean surfaces was done showed no significant difference). Figure 2.4 shows 3-d plots of both the true and PME Poisson intensity surfaces at times $t_1$, $t_{13}$ and $t_{27}$. In addition, Tables 2.2, 2.3 and 2.4 present summary characteristics of the true and the PME intensity surfaces at times $t_1$, $t_{13}$ and $t_{27}$, respectively.

Note that the index assignments for each mixture component are given in the first columns of these tables based on the original way the components were assigned and simulated from the true models, and they may be different from the index assignment when we simulate from the posterior distribution (indices in parentheses). This has no effect on the model fitting procedure other than the fact that we need to take care when we check and interpret the values of the parameters based on the posterior realizations. For example, in Tables 2.2 and 2.3 the assignment is the same as in the truth, whereas, in Table 2.4 Component 1 from the truth is indexed as Component 2 in the posterior simulations, and Component 2 from the truth is indexed as Component 1 (see the values of the estimates to appreciate this index switch). In addition, if we were to start the posterior simulation at different starting values, then this assignment may change, but once we identify the change of the index values, we can interpret the results correctly. Obviously, in real life applications we do not know the...
true index allocation so we do not have to worry about this issue.

However, it should be noted that this index assignment is not the same as the label switching problem that may occur during any iteration of a sampler from the posterior distribution of the mixture parameters. In particular, we used the sppmix R package function check_labels (Chen et al. (2017), https://cran.r-project.org/web/packages/sppmix/index.html) to check for label switching, and the sppmix R package function FixLS_da in order to permute the posterior realizations using an identifiability constraint, if the number of components is more than 7. Otherwise, a decision theoretic approach is used in the case where label switching was detected. The performance of sppmix algorithm for label switching was done (omitted) visually using the trace plots and running means plots before and after label fixing and indeed it works.

2.4.4 Simulation study for the multivariate space-time model

The simulated data for the second model was obtained using the procedure outlined in Section 2.3.5 as follows:

1. To obtain the history $H_T$ we simulate the Hawkes Process with $\mu = 100$, $\alpha = .1$, $\beta = 100$ and $T = 0.1$; i.e., the history over the time window $[0, T)$.

2. Simulate $\gamma(t_k)$ for $t_k, k = 1, \ldots, K$, with $K$ the length of the history, using $\gamma(t_k) \sim U(a = 300, b = 350)$.

3. For each time $t_k \in H_T$, simulate the marks $n(t_k) \sim Poisson(\gamma(t_k))$ truncated from the left at 1 (so that we always have at least one mixture component).

4. Simulate the number of components $m(t_k)$ at each time $t_k \in H_T$ using the evolution parameters $\varphi_\tau = 1$, $\sigma_\tau = 0.1$, $\bar{\mu}_\tau = 0.1$ and $\bar{\sigma}_\tau = 0.5$.

5. (a) Simulate the component probabilities $p(t_k)$ at each time $t_k \in H_T$ using the evolution parameters $\varphi_d = 1$, $\sigma_d = 0.2$, $\bar{\mu}_d = 0.2$ and $\bar{\sigma}_d = 0.2$. 

53
(b) Simulate the component means \( \{\mu_j(t_k)\}_{j=1}^{m(t_k)} \) at each time \( t_k \in \mathcal{H}_T \) using the evolution parameters \( A = 5 \times \text{diag}(2), \Pi = \text{diag}(2), \varphi_{\kappa_1} = 1, \sigma_{\kappa_1} = 0.2, \tilde{\mu}_{\kappa_1} = 0.5, \tilde{\sigma}_{\kappa_1} = 0.5, \varphi_{\kappa_2} = 1, \sigma_{\kappa_2} = 0.2, \tilde{\mu}_{\kappa_2} = 0.5, \tilde{\sigma}_{\kappa_2} = 0.5, \varphi_{\zeta} = 1, \sigma_{\zeta} = 0.2, \tilde{\mu}_{\zeta} = 0.1, \text{ and } \tilde{\sigma}_{\zeta} = 0.5. \)

(c) Simulate the component covariances \( \{\Sigma_j(t_k)\}_{j=1}^{m(t_k)} \) at each time \( t_k \in \mathcal{H}_T \), using the evolution parameters \( \varphi_{\rho_1} = 1, \sigma_{\rho_1} = 0.2, \tilde{\mu}_{\rho_1} = 0.5, \tilde{\sigma}_{\rho_1} = 0.2, \varphi_{\rho_2} = 1, \sigma_{\rho_2} = 0.2, \tilde{\mu}_{\rho_2} = 0.5, \tilde{\sigma}_{\rho_2} = 0.2, \varphi_{\varrho} = 1, \sigma_{\varrho} = 0.2, \tilde{\mu}_{\varrho} = 0.1, \tilde{\sigma}_{\varrho} = 0.2. \)

6. The space-time point pattern is given by \( \varphi_n = \{(s_r(t_k), t_k), \text{ } r = 1, \ldots, n(t_k), \text{ } k = 1, \ldots, K, s_i \in \mathcal{W}, t_k \in \mathcal{T} = [0, T)\} \).

After simulating the data, we implement the MCMC and obtained the following results after running 10,000 iterations with a burn-in period of 2,000 iterations. We obtain the Poisson intensity surfaces based on the posterior means of the parameters at all times \( t_k, k = 1, \ldots, 27 \); however, we only provide details for times \( t_1, t_{13} \) and \( t_{27} \).

2.4.5 Recovering the temporal process

The temporal process part of the DSTPP is governed by the 1-d Hawkes process with intensity function given in equation (2.6) of Section 2.2.1. In Table 2.5, we obtain summary results for the Hawkes process parameters. Figure 2.5 presents plots of the true intensity of the Hawkes process along with the estimated intensity based on the posterior means of the parameters. We note once again that all true parameter values are within the 95% credible sets and the posterior means for the parameters are close to the true values. Figure 2.6 shows the trace plots, histograms and the associated ACF for the parameters \( \mu, \alpha \) and \( \beta \).

2.4.6 Recovering the spatial (location) process

The spatial (location) process part of the DSTPP, conditional on the temporal component, is governed by the NHPPP with intensity function given in equation (2.24). Figure 2.7 shows the 2-d plots of both the true and PME Poisson intensity surfaces, along with the point patterns, at times \( t_1, t_{13} \) and \( t_{27} \). Figure 2.8 shows 3-d plots of both the true and PME
2.5 Modeling Earthquake events in California, USA

In this section we illustrate the methodology by using a real dataset from the Southern California Earthquake Center.

2.5.1 The Southern California Earthquake data

The Southern California Earthquake Center maintains and updates yearly catalogs for earthquakes that occur in Southern California from 1923 to the present and they can be obtained at [http://scedc.caltech.edu/eq-catalogs/](http://scedc.caltech.edu/eq-catalogs/). These catalogs consist of records that include the time of occurrence, estimated epicenter (longitude and latitude), and magnitude of the earthquake, along with a plethora of other variables that will not be utilized in this analysis (e.g., the depth of the earthquake can be used as covariate information).

For the purposes of this study, we consider earthquakes from January 1, 2000, to July 8, 2015 of magnitude greater or equal to 2.50 (the choice of this value was dictated by the fact there is need to have sufficient temporal history (time stamps) so as to estimate the Hawkes process parameters) on the Richter scale, over an observation window with range for longitudes $-125.75^\circ$ W to $-113.15^\circ$ W and latitudes $31.55^\circ$ N to $42.95^\circ$ N. To have enough points at each time period that can be used to fit the mixture models for the location processes, we aggregate the data monthly until we have at least 190 locations. The resulting times $t_k$, $k = 1, \ldots, 44$, represent the average times of occurrences of the time events that have been aggregated within each of the 44 bins.

Figure 2.9 shows a plot of the number of events at each of the aggregated times $t_k$, for $k = 1, \ldots, 44$. Notice the high spike that corresponds to the month of March-April.
2010 is associated with a major earthquake event that occurred near Baja California, on April 4 (about time period \( t_{28} = 123.9707 \)). The magnitude 7.2 earthquake that jolted northern Mexico and Southern California was the first big earthquake to occur on this particular fault system since 1892. Centered on Mexico’s Laguna Salada Fault in Baja California (35.1° North latitude, 115.3° West longitude), the earthquake was triggered by the same processes that drive temblors on the San Andreas Fault, which runs all the way from Southern California to north of San Francisco. In order to visualize the approximate epicenter of the event location, see any of the plots in Figure 2.12 (near the South-eastern most part of California, USA).

This Baja California earthquake killed at least one person but did relatively little damage, because the epicenter was in a lightly populated area 38 miles (60 kilometers) south-southeast of Mexicali, Mexico. Small earthquakes can sometimes precede larger temblors, because the little ones can cause a fault to slip until it reaches a point where a major slip releases a “Big One.” In fact, a magnitude 4.4 temblor had rocked the same region in Baja California about a day before April 4, and there had been other rumblings in the region. More precisely, a separate magnitude 4.4 temblor had jolted the Los Angeles Basin on March 16, centered near the northern end of the Elsinore Fault system. We would like to assess whether or not such smaller magnitude earthquakes can be used to predict subsequent, large magnitude earthquakes.

In order to appreciate the behavior of the space-time process about this major earthquake event, we study times \( t_{26} = 118.149, t_{27} = 120.9912, t_{28} = 123.9707, \) and \( t_{29} = 125 \). That is, we will investigate more closely two time periods before the main earthquake event at \( t_{28} \) and the immediate time after, in an effort to understand the underlying mechanism that led to this devastating earthquake, as well as, the number of potential aftershocks. We also present results at the first and last time periods, e.g., \( t_1 = 2.0237 \) and \( t_{44} = 181.5083 \).
2.5.2 Results based on the marked space-time NHPPP model

The following results were obtained after running 10,000 iterations with a burn-in period of 2,000 iterations. First, we present the results for the temporal (Hawkes process) part and then for parameters of the PME Poisson intensity surfaces.

Recovering the temporal process

Table 2.9 shows summaries of the Hawkes process parameters (i.e., the posterior means and 95% credible sets). Figure 2.10 presents the estimated intensity of the Hawkes process based on the posterior means of the model parameters $\mu$, $\alpha$, and $\beta$, along with the times of occurrence of events (i.e., the temporal point pattern) shown at the bottom of the figure in red segments. Alternatively, we could plot the average of the intensity function based on the posterior realization of the Hawkes model parameters, however, as we found out the functions are not significantly different. Note first how the Hawkes process receives a large number of events about $t_{28} = 123.9707$, thus capturing first the major earthquake event at Baja along with a plethora of aftershocks events that keep the intensity of the process at high levels. After about time 135, the effects of the shocks dissipate and the process returns back to the levels it was before the major earthquake.

Figure 2.11 shows the trace plots, histograms and the associated ACF’s, with the dashed red lines corresponding to the posterior means of the model parameters. Clearly, the chains indicate convergence, and this is also confirmed by plotting the running means (plots not included).

Recovering the spatial (location) process

Figures 2.12 and 2.13 show the 2-d plots (along with the point patterns and the component means) and the corresponding 3-d plots of the PME Poisson intensity surfaces, respectively, at times $t_1$, $t_{26}$, $t_{27}$, $t_{28}$, $t_{29}$, and $t_{44}$.

Based on the 3d figures (plots b-e), we notice the spikes in the intensity of the locations
of the earthquake events just before, during and after time $t_{28}$, indicating a large cluster of events in the Laguna Salada Fault in Baja California, as anticipated.

Posterior characteristics for the model parameters over these times are obtained in tables 2.10, 2.11, 2.12, 2.13, 2.14, and 2.15 respectively. The 2d and 3d plots are based on the values of these tables, and they are used to create and display the PME Poisson intensity surfaces.

Concentrating on the time periods about time $t_{28}$, we consider the mixture deconvolution in the tables and notice that we use 7, 7, 17 and 6, as the number of components at times $t_{26}$, $t_{27}$, $t_{28}$, and $t_{29}$. This increase in the required number of components used to describe the location process, is yet another indication that at time $t_{28}$ there is an unusual cluster of events, vis-a-vis, a large number of events before April 4, the large earthquake at April 4, and a large number of events after April 4 (aftershocks).

### 2.5.3 Results based on the multivariate space-time model

The following results were obtained after running 10,000 iterations with a burn-in period of 2,000 iterations. First, we present the results for the temporal (Hawkes process) part and then for parameters of the PME Poisson intensity surfaces.

**Recovering the temporal process**

Table 2.16 shows summaries of the Hawkes process parameters (i.e., the posterior means and 95% credible sets), whereas, Figure 2.14 presents the estimated intensity of the Hawkes process based on the posterior means of the model parameters $\mu$, $\alpha$ and $\beta$, along with the times of occurrence of events (i.e., the temporal point pattern) shown at the bottom of the figure in red segments.

Figure 2.15 shows the trace plots, histograms and the associated ACF’s, with the dashed red lines corresponding to the posterior means of the Model parameters. A visual assessment of the trace plots, indicated the chains converge, and this is also confirmed by plotting the running means (plots not included).
Note that the fit and interpretation of the results is similar to the results on Model 1. However, we note that we used a slightly different prior on the parameter $\beta$. That is, instead of an exponential distribution we considered a Gamma distribution with shape parameter $\alpha = 2$ and scale parameter $b_\beta$ selected based on the observed time events $t_k, k = 1, \ldots, 44$ (empirical Bayes prior).

**Recovering the spatial (location) process**

Figures 2.16 and 2.17 show the 2-d plots (along with the point patterns and the component means) and the corresponding 3-d plots of the PME Poisson intensity surfaces, respectively, at times $t_1, t_26, t_27, t_28, t_29, \text{ and } t_{44}$. The interpretation of the results is again similar to those in Model 1, and consistent with the behavior of the space-time process about the time of the major earthquake event at time $t_{28}$. In addition, the posterior characteristics for the model parameters over these times are displayed in Tables 2.17, 2.18, 2.19, 2.20, 2.21, and 2.22, respectively.

### 2.6 Discussion

We presented a hierarchical Bayesian framework for modeling the intensity function of the DSTPP using finite mixtures for the location processes and the one dimensional Hawkes Process for the temporal events, via conditioning (i.e., $\lambda(s, t) = \lambda \times \lambda(t) \times \lambda(s \mid t)$). The conditioning considered in this chapter is on the time stamps (occurrence times of the spatial events). It is evident from the results above from both the simulated data and the application to earthquake data, that the models presented herein provide an appealing framework to model and capture the dynamics of a spatio-temporal point process.

The methodology in this chapter provides a framework that can be thought of an extension of Poisson point process models that can include a time stamp (discrete or continuous, in this exposition continuous). Using the Hawkes process, which is itself a Poisson Point process, allows us to model and obtain the time stamps. By using the Hawkes process we
adhere to the natural ordering of time and furthermore we can obtain an analytical solution of $\Lambda(T)$. The latter is typically approximated using numerical methods and thus the results obtained in this chapter are exact and no approximations in the model likelihood are required.

By extending the conditioning approach of [Micheas (2011)](#) we are able to model the space-time Poison point process intensity $\lambda(s,t)$ that allows replications by treating the times of events as a mark. Furthermore, we considered another approach in the multivariate space time model, which equally accommodates replications, with a major difference being that we treat the whole point pattern at a given time as a mark.

It is worth noting that the dimension of the parameter space changes across time because of the varying number of mixture components, which increases model complexity. Evolving both the truncated-Poisson parameter for the number of components and the mean for the Dirichlet parameter helps us in circumventing the varying dimension. In contrast to using infinite mixtures, the model presented provides summary statistics for different features of the intensity surface (mixture deconvolution), thus providing a means by which inference can be made about a specific feature that may be of interest. For example, the mean of a specific mixture component would be of interest in certain applications, such as, finding epicenters of major earthquakes, crime hot-spots in a city, allocation of emergency centers for ambulance demand, or centers for disease outbreaks, to name a few. Therefore, mixture deconvolution can help policy makers to efficiently allocate resources when faced with potential loss of life and property. Ultimately, we have been able to capture the spatial variation and the temporal clustering of earthquakes that occur in Southern California.
Appendix: Full Conditional Distributions

2.A Full-Conditional Distributions

In this section we present the full-conditional distributions for both the marked space-time NHPPP model and the multivariate space-time model, based on the full posterior distributions (2.38) and (2.43), respectively. The full conditional distributions of the non-dynamic parameters of the Hawkes process parameters $\mu, \alpha$, and $\beta$ and parameters $\lambda$ and $A$ are presented first. Next, we present the full-conditional distributions for the dynamically evolving parameters $\gamma(t_k), \tau(t_k), m(t_k), d(t_k), p(t_k), \mu_j(t_k), \Sigma_j(t_k), \psi(t_k), \kappa_1(t_k), \kappa_2(t_k), \zeta(t_k), \rho_1(t_k), \rho_2(t_k), \varrho(t_k)$ for $k = 1, \ldots, K$, and finally, full-conditional distributions for the hyper-parameters $\sigma_\gamma^2, \sigma_\tau^2, \sigma_{\kappa_1}^2, \sigma_{\kappa_2}^2, \sigma_\zeta^2, \sigma_\rho_1^2, \sigma_\rho_2^2, \sigma_\varrho^2$.

2.A.1 Updating the non-dynamic parameters

The proposal distributions for the Hawkes process parameters $\mu, \alpha$ and $\beta$ are given by $\mu^{\text{prop}} \sim N(\mu, \sigma_\mu^2)$, truncated in $(0, \infty)$, $\alpha^{\text{prop}} \sim N(\alpha, \sigma_\alpha^2)$ truncated in $(0,1)$, and $\beta^{\text{prop}} \sim N(\beta, \sigma_\beta^2)$, truncated in $(0, \infty)$. The Metropolis-Hasting ratios for updating $\mu, \alpha$ and $\beta$ are given by

$$MH_\mu = \frac{\pi(\mu^{\text{prop}}, \alpha, \beta|\cdot)}{\pi(\mu, \alpha, \beta|\cdot)} \times \frac{p(\mu^{\text{prop}})}{p(\mu)},$$

$$MH_\alpha = \frac{\pi(\mu, \alpha^{\text{prop}}, \beta|\cdot)}{\pi(\mu, \alpha, \beta|\cdot)} \times \frac{p(\alpha^{\text{prop}})}{p(\alpha)},$$

$$MH_\beta = \frac{\pi(\mu, \alpha, \beta^{\text{prop}}|\cdot)}{\pi(\mu, \alpha, \beta|\cdot)} \times \frac{p(\beta^{\text{prop}})}{p(\beta)},$$

where $\pi(\mu, \alpha, \beta|\cdot)$ is given by (2.42) or (2.44) depending on the model used.

The full-conditional distribution for $\lambda$ is simply $\lambda \sim G(n + a_\lambda, (1 + 1/b_\lambda)^{-1})$, whereas, the covariance matrix for the component means $A$ has a full conditional distribution given by

$$A|\cdot \sim IWishart_2 \left(2a_2 + \sum_{k=1}^K m(t_k), 2Q + \sum_{k=1}^K \sum_{j=1} m(t_k)(\mu_j(t_k) - \Psi(t_k))(\mu_j(t_k) - \Psi(t_k))^T \right).$$
2.A.2 Updating the mean of number of components $\tau(t_k)$

The full-conditional distribution for the mean number of components $\tau(t_k)$ at time $t_k$ is given by

$$
\pi(\tau(t_1) | \cdot) \propto \frac{e^{-\tau(t_1)}(\tau(t_1))^{m(t_1)}}{m(t_k)!S(\tau(t_k), k_1, k_2)} \times \exp \left\{ -\frac{\sigma_\tau^2 + \varphi_\tau^2 \sigma_\tau^2}{2\sigma_\tau^2\sigma_\tau^2} \left( \log(\tau(t_1)) - \frac{\bar{\mu}_\tau \sigma_\tau^2 + \varphi_\tau \sigma_\tau^2 \log(\tau(t_2))}{\sigma_\tau^2 + \varphi_\tau^2 \sigma_\tau^2} \right)^2 \right\},
$$

for $k = 1$, by

$$
\pi(\tau(t_k) | \cdot) \propto \frac{e^{-\tau(t_k)}(\tau(t_k))^{m(t_k)}}{m(t_k)!S(\tau(t_k), k_1, k_2)} \times \exp \left\{ -\frac{1 + \varphi_\tau^2}{2\sigma_\tau^2} \left( \log(\tau(t_k)) - \frac{\varphi_\tau \log(\tau(t_k-1)) + \log(\tau(t_{k+1}))}{1 + \varphi_\tau^2} \right)^2 \right\},
$$

$k = 2, \ldots, K - 1$, and by

$$
\pi(\tau(t_K) | \cdot) \propto \frac{e^{-\tau(t_K)}(\tau(t_K))^{m(t_k)}}{m(t_k)!S(\tau(t_k), k_1, k_2)} \times \exp \left\{ -\frac{1}{2\sigma_\tau^2} (\log(\tau(t_K)) - \varphi_\tau \log(\tau(t_{K-1})))^2 \right\},
$$

for $k = K$, where $S(\tau(t_k), k_1, k_2)$ is the normalizing constant for the truncated Poisson distribution truncated between $k_1$ and $k_2$ given by

$$
S(\tau(t_k), k_1, k_2) = \sum_{m=k_1}^{k_2} \frac{(\tau(t_k))^m e^{-\tau(t_k)}}{m!}.
$$

In addition, the related hyper-parameters to $\tau(t_k)$ have full conditional distributions given by

$$
\sigma_\tau^2 | \cdot \sim IG \left( \frac{K + 2a_\tau - 1}{2}, \left[ \frac{1}{2} \sum_{k=2}^{K} (\log(\tau(t_k)) - \varphi_\tau \log(\tau(t_{k-1})))^2 + 1/b_\tau \right]^{-1} \right),
$$

(2.52)
with \( \tilde{\sigma}_\tau^2 \cdot \sim IG \left( \tilde{A}_\tau, \tilde{B}_\tau \right) \), where

\[
\tilde{A}_\tau = \bar{a}_\tau + \frac{1}{2}, \quad \tilde{B}_\tau = \left[ \frac{1}{2} (\log(t_1)) - \bar{\mu}_\tau \right]^2 + 1/\bar{b}_\tau \]

(2.53)

and \( \tilde{\mu}_\tau \cdot \sim N(M_{\tilde{\mu}_\tau}, V_{\tilde{\mu}_\tau}) \), where

\[
V_{\tilde{\mu}_\tau} = \frac{\tilde{\sigma}_\tau^2 \sigma^2_\tau}{\sigma^2_\tau + \sigma^2_\tau},
\]

(2.54)

and

\[
M_{\tilde{\mu}_\tau} = V_{\tilde{\mu}_\tau} \log(t_1).
\]

(2.55)

Note that when we fit the models, we set \( \varphi_\tau = 1 \) everywhere above. In the case where we use \( \varphi_\tau \sim \pi(\varphi_\tau) \propto 1 \), the full conditional for \( \varphi_\tau \) becomes

\[
\varphi_\tau \mid \cdot \sim N \left( \frac{\sum_{k=2}^{K} \log(\tau(t_k)) \log(\tau(t_{k-1}))}{K \sum_{k=2}^{K} \log(\tau(t_{k-1}))^2}, \frac{\sigma^2_\tau}{\sum_{k=2}^{K} \log(\tau(t_{k-1}))^2} \right).
\]

(2.56)

Finally, we note that the M-H ratio for updating \( \tau(t_k) \) is given by

\[
H^k_\tau = \left( \frac{\tau^{(prop)}(t_k)}{\tau(t_k)} \right)^{m(t_k)} \frac{1 - e^{-\tau(t_k)}}{1 - e^{-\tau^{(prop)}(t_k)}} e^{-(\tau^{(prop)}(t_k) - \tau(t_k))},
\]

(2.57)

with the proposal distributions being log-normals with the following parameter means and variances; for \( k = 1 \) we let

\[
\text{mean} = \frac{\tilde{\mu}_\tau \sigma^2_\tau + \varphi_\tau \tilde{\sigma}_\tau^2 \log(\tau(t_2))}{\sigma^2_\tau + \varphi^2_\tau \tilde{\sigma}_\tau^2}, \quad \text{variance} = \frac{\sigma^2_\tau \tilde{\sigma}_\tau^2}{\sigma^2_\tau + \varphi^2_\tau \tilde{\sigma}_\tau^2},
\]

(2.58)

for \( k = 2, \ldots, K - 1 \),

\[
\text{mean} = \frac{\varphi_\tau (\log(\tau(t_{k-1})) + \log(\tau(t_{k+1}))))}{1 + \varphi^2_\tau}, \quad \text{variance} = \frac{\sigma^2_\tau}{1 + \varphi^2_\tau},
\]

(2.59)
and for $k = K$,

$$\text{mean} = \varphi \log(\tau(t_{K-1})), \quad \text{variance} = \sigma^2. \quad (2.60)$$

### 2.A.3 Updating the mean of Dirichlet parameter $\mu_d(t_k)$

The full-conditional distribution for the means $\mu_d(t_k)$ of the Dirichlet parameters at time $t_k$; i.e., the mean of the $j^{th}$ component Dirichlet parameter $d_j(t_k)$, are derived as follows; for $k = 1$, we have

$$\mu_d(t_1)|· \sim N(M_{d_1}, V_{d_1}),$$

where

$$V_{d_1} = \frac{\tilde{\sigma}_d^2 \sigma_d^*^2}{m(t_1) \tilde{\sigma}_d^2 \sigma_d^2 + \sigma_d^*^2 \sigma_d^2 + \varphi_d^2 \sigma_d^*^2 \tilde{\sigma}_d^2}, \quad (2.61)$$

and

$$M_d = \frac{\tilde{\mu}_d \tilde{\sigma}_d^2 \sigma_d^2 + \varphi_d \mu_d(t_2) \tilde{\sigma}_d^2 \sigma_d^2 + \tilde{\sigma}_d^2 \sum_{j=1}^{m(t_k)} \log(d_j(t_1))}{m(t_1) \tilde{\sigma}_d^2 \sigma_d^2 + \sigma_d^*^2 \sigma_d^2 + \varphi_d^2 \sigma_d^*^2 \tilde{\sigma}_d^2} \quad (2.62)$$

For $k = 2, \ldots, K - 1$, we have $\mu_d(t_k)|· \sim N(M_{d_k}, V_{d_k})$, where

$$V_{d_k} = \frac{\sigma_d^2 \sigma_d^*^2}{m(t_k) \sigma_d^2 + \sigma_d^*^2 + \varphi_d^2 \sigma_d^*^2}, \quad (2.63)$$

and

$$M_d = \frac{\varphi_d \sigma_d^*^2 + \varphi_d \sigma_d^2 \mu_d(t_{k-1}) + \sigma_d^2 \mu_d(t_{k+1}) \sum_{j=1}^{m(t_k)} \log(d_j(t_k))}{m(t_k) \sigma_d^2 + \sigma_d^*^2 + \varphi_d^2 \sigma_d^*^2}, \quad (2.64)$$

and for $k = K$, $\mu_d(t_K)|· \sim N(M_{d_K}, V_{d_K})$, where

$$V_{d_K} = \frac{\sigma_d^2 \sigma_d^*^2}{\sigma_d^*^2 + m(t_K) \sigma_d^2}, \quad (2.65)$$

and

$$M_d = \frac{\varphi_d \sigma_d^*^2 \mu_d(t_{K-1}) + \sigma_d^2 \sum_{j=1}^{m(t_K)} \log(d_j(t_K))}{\sigma_d^*^2 + m(t_K) \sigma_d^2}. \quad (2.66)$$
The related hyper-parameters to $\mu_d(t_k)$ have full conditional distributions given by $\sigma_d^2 \sim IG(A_d, B_d)$, where

$$A_d = \frac{1}{2}[K + 2a_d - 1], \quad B_d = \left[\frac{1}{2} \sum_{k=2}^{K} [\log(\mu_d(t_k)) - \varphi_d \log(\mu_d(t_{k-1}))]^2 + 1/b_d\right]^{-1}, \quad (2.67)$$

$\sigma_d^{*2} \sim IG(A^*_d, B^*_d)$, where

$$A^*_d = \frac{1}{2}\left[\sum_{k=1}^{K} m(t_k) + 2a^*_d - 1\right], \quad B^*_d = \left[\frac{1}{2} \sum_{k=1}^{K} \sum_{j=1}^{m(t_k)} (\log(d_j(t_k)) - \mu_d(t_k))^2 + 1/b^*_d\right]^{-1}, \quad (2.68)$$

$\tilde{\sigma}_d^2 \sim IG(\tilde{A}_d, \tilde{B}_d)$, where

$$\tilde{A}_d = \frac{1}{2}[2\tilde{a}_d + 1], \quad \tilde{B}_d = \left[\frac{1}{2}(\mu_d(t_1) - \tilde{\mu}_d)^2 + 1/b_d\right]^{-1}, \quad (2.69)$$

and $\tilde{\mu}_d \sim N(M_{\tilde{\mu}_d}, V_{\tilde{\mu}_d})$, where

$$V_{\tilde{\mu}_d} = \frac{\tilde{\sigma}_d^2 \sigma_d^2}{\sigma_d^2 + \sigma_d^{*2}}, \quad (2.70)$$

and

$$M_{\tilde{\mu}_d} = V_{\tilde{\mu}_d} \mu_d(t_1). \quad (2.71)$$

Note that when we fit the models, we set $\varphi_d = 1$ everywhere above. In the case where we use $\varphi_d \sim \pi(\varphi_d) \propto 1$, the full conditional for $\varphi_d$ becomes

$$\varphi_d \sim N \left( \frac{\sum_{k=2}^{K} \log(d(t_k)) \log(d(t_{k-1}))}{\sum_{k=2}^{K} (\log(d(t_{k-1})))^2}, \frac{\sigma_d^2}{\sum_{k=2}^{K} (\log(d(t_{k-1})))^2} \right). \quad (2.72)$$

### 2.A.4 Updating the Dirichlet parameters $d(t_k)$

The full-conditional distribution of the $j^{th}$ component of the Dirichlet parameter $d_j(t_k)$ at time $t_k$ is given by

$$\log d_j(t_k) \sim N(\mu_d(t_k), \sigma_d^2), \quad (2.73)$$
for $j = 1, \ldots, m(t_k)$, and $k = 1, \ldots, K$.

2.A.5 Updating the membership indicator vector $z_r(t_k)$

The full-conditional distribution for the membership indicator vector $z_r(t_k)$ is given by

\[ z_r(t_k) \mid \cdot \sim \text{Multinomial}(1, q_{r,1}(t_k), \ldots, q_{r,m(t_k)}(t_k)), \quad (2.74) \]

where

\[ q_{r,j}(t_k) = \frac{p_j(t_k) \varphi_{j,t_k}(s_r(t_k)|\mu_j(t_k), \Sigma_j(t_k))}{\sum_{i=1}^{m(t_k)} p_i(t_k) \varphi_{i,t_k}(s_r(t_k)|\mu_i(t_k), \Sigma_i(t_k))}, \quad (2.75) \]

for $j = 1, \ldots, m(t_k)$, $r = 1, \ldots, n(t_k)$, and $k = 1, \ldots, K$.

2.A.6 Updating the component probabilities $p(t_k)$

The full-conditional distribution for the component probabilities vector $p(t_k)$ at time $t_k$ is given by

\[ p(t_k) \mid \cdot \sim \text{Dirichlet}(d_1(t_k) + n_1(t_k), \ldots, d_{m(t_k)}(t_k) + n_{m(t_k)}(t_k)), \quad (2.76) \]

where $n_j(t_k) = \sum_{r=1}^{n(t_k)} z_{r,j}(t_k)$, $j = 1, \ldots, m(t_k)$, represents the number of observations assigned to the $j^{th}$ component at time $t_k$, for $k = 1, \ldots, K$.

2.A.7 Updating the $j^{th}$ component mean $\mu_j(t_k)$

The full-conditional distribution for the $j^{th}$ component mean at time $t_k$ is given by

\[ \mu_j(t_k) \mid \cdot \sim N_2(M_j(t_k), V_j(t_k)), \quad (2.77) \]

where

\[ M_j(t_k) = V_j(t_k) \left( n_j(t_k) \Sigma_j^{-1}(t_k) s_j(t_k) + A^{-1} \Psi(t_k) \right), \]
and

\[ V_j(t_k) = (n_j(t_k)\Sigma_j^{-1}(t_k) + A^{-1})^{-1}, \]

with \( \bar{s}_j(t_k) = \frac{1}{n_j(t_k)} \sum_{r=1}^{n(t_k)} z_{r,j}(t_k)s_r(t_k) \), for \( j = 1, \ldots, m(t_k) \), and \( k = 1, \ldots, K \).

2.A.8 Updating the \( j \)th component covariance matrix \( \Sigma_j(t_k) \)

The full-conditional distribution for the \( j \)th component covariance matrix at time \( t_k \), is given by

\[ \Sigma_j(t_k) \mid \cdot \sim \text{iWishart}_2(2a_1 + n_j(t_k)), Q_j(t_k)), \tag{2.78} \]

where

\[ Q_j(t_k) = \left[ B^{-1}(t_k) + \sum_{r=1}^{n(t_k)} z_{r,j}(t_k)(s_r(t_k) - \mu_j(t_k))(s_r(t_k) - \mu_j(t_k))^T \right]^{-1}, \]

for \( j = 1, \ldots, m(t_k) \), and \( k = 1, \ldots, K \).

2.A.9 Updating the mean of the components means \( \psi(t_k) \)

The full-conditional distribution for the mean vector \( \psi(t_k) \) of component means at time \( t_k \) is computed as follows. For \( k = 1 \), we have \( \psi(t_1) \mid \cdot \sim N_2(M_{\psi_1}, V_{\psi_1}) \), where

\[ V_{\psi_1} = \left( \Pi^T \Delta^{-1}(t_2)\Pi + m(t_1)A^{-1} + \bar{\Sigma}^{-1} \right)^{-1}, \tag{2.79} \]

\[ M_{\psi_1} = V_{\psi_1} \left( \bar{\Sigma}^{-1}\bar{\psi} + \Pi^T \Delta^{-1}(t_2)\psi(t_2) + A^{-1} \sum_{j=1}^{m(t_1)} \mu(t_1) \right), \tag{2.80} \]

for \( k = 2, \ldots, K - 1 \), we have \( \psi(t_k) \mid \cdot \sim N_2(M_{\psi_k}, V_{\psi_k}) \), where

\[ V_{\psi_k} = (\Pi^T \Delta^{-1}(t_{k+1})\Pi + m(t_k)A^{-1} + \Delta^{-1}(t_{k-1}))^{-1}, \tag{2.81} \]

\[ M_{\psi_k} = V_{\psi_k} \left( \Delta^{-1}(t_{k-1})\Pi\psi(t_{k-1}) + \Delta^{-1}(t_{k+1})\Pi\psi(t_{k+1}) + A^{-1} \sum_{j=1}^{m(t_k)} \mu(t_k) \right), \tag{2.82} \]
and for \( k = K \), we have \( \psi(t_K) | \sim N_2(M_{\psi_K}, V_{\psi_K}) \), where

\[
V_{\psi_K} = (m(t_K)A^{-1} + \Delta^{-1}(t_{K-1}))^{-1},
\]

\[
M_{\psi_K} = V_{\psi_K}(\Delta^{-1}(t_{K-1})\Pi(t_{K-1}) + A^{-1}\sum_{j=1}^{m(t_K)} \mu(t_K)).
\]

Note that when we fit the models, we set \( \Pi = I_2 \), everywhere above. In the case where we use \( \Pi \sim \pi(\Pi) \propto 1 \), the full conditional for \( \Pi \) becomes

\[
\pi(\Pi | \cdot) \propto \exp \left\{ -\frac{1}{2} \sum_{k=2}^{K} (\psi(t_k) - \Pi(t_{k-1})^T\Delta^{-1}(t_k)(\psi(t_k) - \Pi(t_{k-1})) \right\},
\]

where \( \Pi = \text{diag}(\pi_1, \pi_2) \). The proposal distribution for \( \pi_1 \) is normal with mean the current value and standard deviation \( \sigma_{\pi_1} \) (i.e., \( \pi_1^{(\text{prop})} \sim N(\pi_1, \sigma_{\pi_1}^2) \)) and proposal distribution for \( \pi_2 \) is also normal with mean the current value and standard deviation \( \sigma_{\pi_2} \), i.e., \( \pi_2^{(\text{prop})} \sim N(\pi_2, \sigma_{\pi_2}^2) \).

Then the M-H ratio for sampling \( \Pi \) is given by

\[
H_{\Pi} = \exp \left\{ -\frac{1}{2} \sum_{k=2}^{K} (\psi(t_k) - \Pi(t_{k-1})^T\Delta^{-1}(t_k)(\psi(t_k) - \Pi(t_{k-1})) \right\} \cdot \frac{\exp \left\{ -\frac{1}{2} \sum_{k=2}^{K} (\psi(t_k) - \Pi(t_{k-1})^T\Delta^{-1}(t_k)(\psi(t_k) - \Pi(t_{k-1})) \right\}}{\exp \left\{ -\frac{1}{2} \sum_{k=2}^{K} (\psi(t_k) - \Pi(t_{k-1})^T\Delta^{-1}(t_k)(\psi(t_k) - \Pi(t_{k-1})) \right\}},
\]

where \( \Pi^{(\text{prop})}(t_k) = \text{diag}(\pi_1^{(\text{prop})}, \pi_2^{(\text{prop})}) \).

2.A.10 Updating the covariance matrix \( \Delta(t_k) \) of the component means \( \psi(t_k) \)

The full-conditional distribution for the covariance matrix \( \Delta(t_k) \) of the mean of the component means at time \( t_k \). Recall that \( \Delta(t_k) = K(t_k)\Gamma(t_k)K(t_k)^T \), with \( K(t_k) = \text{diag}(\kappa_1(t_k), \kappa_2(t_k)) \) and \( \Gamma(t_k) = \begin{pmatrix} 1 & 0 \\ \zeta(t_k) & 1 \end{pmatrix} \). The full-conditional distributions of the elements of \( \Delta(t_k) \) are given below.
For $k = 1$, the full-conditional distribution of $\kappa_1(t_k)$ is given by

$$
\pi(\kappa_1(t_1) | \cdot) \propto | \Delta(t_1) |^{-\frac{1}{2}} \times \exp\left\{ \left( -\frac{1}{2} \left( (\psi(t_1) - \bar{\mu}_\psi)^T \Delta^{-1}(t_1) (\psi(t_1) - \bar{\mu}_\psi) \right) \right\} \times \exp\left\{ -\frac{\sigma_{\kappa_1}^2 + \varphi_{\kappa_1} \bar{\sigma}_{\kappa_1}^2}{2 \sigma_{\kappa_1}^2 \bar{\sigma}_{\kappa_1}^2} \left( \log(\kappa_1(t_1)) - \frac{\bar{\mu}_{\kappa_1} \sigma_{\kappa_1}^2 + \varphi_{\kappa_1} \bar{\sigma}_{\kappa_1}^2 \log(\kappa_1(t_2))}{\sigma_{\kappa_1}^2 + \varphi_{\kappa_1} \bar{\sigma}_{\kappa_1}^2} \right)^2 \right\},
$$

(2.87)

for $k = 2, \ldots, K - 1$,

$$
\pi(\kappa_1(t_k) | \cdot) \propto | \Delta(t_k) |^{-\frac{1}{2}} \times \exp\left\{ \left( -\frac{1}{2} \left( (\psi(t_k) - \Pi \psi(t_{k-1}))^T \Delta^{-1}(t_k) (\psi(t_k) - \Pi \psi(t_{k-1})) \right) \right\} \times \exp\left\{ -\frac{1 + \varphi_{\kappa_1}^2}{2 \sigma_{\kappa_1}^2} \left( \log(\kappa_1(t_k)) - \frac{\varphi_{\kappa_1} (\log(\kappa_1(t_{k-1})) + \log(\kappa_1(t_{k+1}))))}{1 + \varphi_{\kappa_1}^2} \right)^2 \right\},
$$

(2.88)

and for $k = K$,

$$
\pi(\kappa_1(t_K) | \cdot) \propto | \Delta(t_K) |^{-\frac{1}{2}} \times \exp\left\{ -\frac{1}{2} \left( \psi(t_K) - \Pi \psi(t_{K-1}))^T \Delta^{-1}(t_K) (\psi(t_K) - \Pi \psi(t_{K-1})) \right) \right\} \times \exp\left\{ -\frac{1}{2 \sigma_{\kappa_1}^2} \left( \log(\kappa_1(t_K)) - \varphi_{\kappa_1} \log(\kappa_1(t_{K-1})) \right)^2 \right\}.
$$

(2.89)

We note that the M-H ratio for updating $\kappa_1(t_k)$ is given by

$$
H_{\kappa_1}^k = \frac{| \Delta^{(prop)}(t_k) |^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} \left( (\psi(t_k) - \Pi \psi(t_{k-1}))^T (\Delta^{(prop)}(t_k))^{-1} (\psi(t_k) - \Pi \psi(t_{k-1})) \right) \right\}}{| \Delta(t_k) |^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} \left( (\psi(t_k) - \Pi \psi(t_{k-1}))^T \Delta^{-1}(t_k) (\psi(t_k) - \Pi \psi(t_{k-1})) \right) \right\}},
$$

(2.90)

where $\Delta^{(prop)}(t_k) = K^{(prop)}(t_k) \Gamma(t_k) \Gamma(t_k)^T K^{(prop)}(t_k)$ with $K^{(prop)}(t_k) = \text{diag}(\kappa_1^{(prop)}(t_k), \kappa_2(t_k))$, and $\Gamma(t_k) = \begin{pmatrix} 1 & 0 \\ \zeta(t_k) & 1 \end{pmatrix}$. The proposal distributions are log-normals with the following parameter means and variances. For $k = 1$, we have

$$
\text{mean} = \frac{\bar{\mu}_{\kappa_1} \sigma_{\kappa_1}^2 + \varphi_{\kappa_1} \bar{\sigma}_{\kappa_1}^2 \log(\kappa_1(t_2))}{\sigma_{\kappa_1}^2 + \varphi_{\kappa_1} \bar{\sigma}_{\kappa_1}^2}, \text{variance} = \frac{\sigma_{\kappa_1}^2 \bar{\sigma}_{\kappa_1}^2}{\sigma_{\kappa_1}^2 + \varphi_{\kappa_1} \bar{\sigma}_{\kappa_1}^2},
$$

(2.91)
for $k = 2, \ldots, K - 1$,
\[
\text{mean} = \frac{\varphi_{\kappa_1} (\log(\kappa_1(t_{k-1})) + \log(\kappa_1(t_{k+1})))}{1 + \varphi_{\kappa_1}^2}, \quad \text{variance} = \frac{\sigma_{\kappa_1}^2}{1 + \varphi_{\kappa_1}^2}, \tag{2.92}
\]
and for $k = K$,
\[
\text{mean} = \varphi_{\kappa_1} \log(\kappa_1(t_{K-1})), \quad \text{variance} = \sigma_{\kappa_1}^2. \tag{2.93}
\]

The full-conditional distribution of $\kappa_{2}(t_k)$ is given next. For $k = 1$ we have
\[
\pi(\kappa_{2}(t_1)|\cdot) \propto |\Delta(t_1)|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} \left( \psi(t_1) - \tilde{\mu}_\psi \right)^T \Delta^{-1}(t_1)(\psi(t_1) - \tilde{\mu}_\psi) \right\} \times \exp \left\{ -\frac{\sigma_{\kappa_2}^2 + \varphi_{\kappa_2}^2 \tilde{\sigma}_{\kappa_2}^2}{2\sigma_{\kappa_2}^2 \tilde{\sigma}_{\kappa_2}^2} \left( \log(\kappa_{2}(t_1)) - \frac{\tilde{\mu}_{\kappa_2} \sigma_{\kappa_2}^2 + \varphi_{\kappa_2} \tilde{\sigma}_{\kappa_2}^2 \log(\kappa_{2}(t_2))}{\sigma_{\kappa_2}^2 + \varphi_{\kappa_2}^2 \tilde{\sigma}_{\kappa_2}^2} \right)^2 \right\}, \tag{2.94}
\]
for $k = 2, \ldots, K - 1$,
\[
\pi(\kappa_{2}(t_k)|\cdot) \propto |\Delta(t_k)|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} \left( \psi(t_k) - \Pi \psi(t_{k-1}) \right)^T \Delta^{-1}(t_k)(\psi(t_k) - \Pi \psi(t_{k-1})) \right\} \times \exp \left\{ -\frac{1 + \varphi_{\kappa_2}^2}{2\sigma_{\kappa_2}^2} \left( \log(\kappa_{2}(t_k)) - \frac{\varphi_{\kappa_2} (\log(\kappa_{2}(t_{k-1}))) + \log(\kappa_{2}(t_{k+1})))}{1 + \varphi_{\kappa_2}^2} \right)^2 \right\}, \tag{2.95}
\]
and for $k = K$,
\[
\pi(\kappa_{2}(t_K)|\cdot) \propto |\Delta(t_K)|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} \left( \psi(t_K) - \Pi \psi(t_{K-1}) \right)^T \Delta^{-1}(t_K)(\psi(t_K) - \Pi \psi(t_{K-1})) \right\} \times \exp \left\{ -\frac{1}{2\sigma_{\kappa_2}^2} (\log(\kappa_{2}(t_K)) - \varphi_{\kappa_2} \log(\kappa_{2}(t_{K-1})))^2 \right\}. \tag{2.96}
\]
We note that the M-H ratio for updating \( \kappa_2(t_k) \) is given by

\[
H_{\kappa_2}^k = \frac{|\Delta^{(\text{prop})}(t_k)|^{\frac{1}{2}} \exp\{-\frac{1}{2}(\psi(t_k) - \Pi \psi(t_{k-1}))^T \Delta^{(\text{prop})}(t_k)^{-1}(\psi(t_k) - \Pi \psi(t_{k-1})))\}}{|\Delta(t_k)|^{\frac{1}{2}} \exp\{-\frac{1}{2}(\psi(t_k) - \Pi \psi(t_{k-1}))^T \Delta^{-1}(t_k)(\psi(t_k) - \Pi \psi(t_{k-1})))\}},
\]

(2.97)

where \( \Delta^{(\text{prop})}(t_k) = K^{(\text{prop})}(t_k) \Gamma(t_k) K^{(\text{prop})}(t_{k}) \), \( K^{(\text{prop})}(t_k) = \text{diag}(\kappa_1(t_k), \kappa_2^{(\text{prop})}(t_k)) \), and \( \Gamma(t_k) = \begin{pmatrix} 1 & 0 \\ \zeta(t_k) & 1 \end{pmatrix} \). The proposal distributions are log-normals with the following parameter means and variances. For \( k = 1 \), we have

\[
\text{mean} = \frac{\bar{\mu}_{\kappa_2} \sigma_{\kappa_2}^2 + \varphi_{\kappa_2} \bar{\sigma}_{\kappa_2}^2 \log(\kappa_2(t_2))}{\sigma_{\kappa_2}^2 + \varphi_{\kappa_2} \bar{\sigma}_{\kappa_2}^2}, \quad \text{variance} = \frac{\sigma_{\kappa_2}^2 \bar{\sigma}_{\kappa_2}^2}{\sigma_{\kappa_2}^2 + \varphi_{\kappa_2} \bar{\sigma}_{\kappa_2}^2},
\]

(2.98)

for \( k = 2, \ldots, K - 1 \),

\[
\text{mean} = \frac{\varphi_{\kappa_2} \log(\kappa_2(t_{k-1})) + \log(\kappa_2(t_{k+1}))}{1 + \varphi_{\kappa_2}^2}, \quad \text{variance} = \frac{\sigma_{\kappa_2}^2}{1 + \varphi_{\kappa_2}^2},
\]

(2.99)

and for \( k = K \),

\[
\text{mean} = \varphi_{\kappa_2} \log(\kappa_2(t_{K-1})), \quad \text{variance} = \sigma_{\kappa_2}^2.
\]

(2.100)

The full-conditional distribution of \( \zeta(t_k) \), for \( k = 1 \) is given by

\[
\pi(\zeta(t_1) | \cdot) \propto |\Delta(t_1)|^{-\frac{1}{2}} \times \exp\left\{ -\frac{1}{2}(\psi(t_k) - \bar{\mu}_\psi)^T \Delta^{-1}(t_1)(\psi(t_k) - \bar{\mu}_\psi) \right\} \times \exp\left\{ -\frac{\sigma_{\zeta}^2 + \varphi_{\zeta}^2 \bar{\sigma}_{\zeta}^2}{2\sigma_{\zeta}^2 \bar{\sigma}_{\zeta}^2} \left( \zeta(t_1) - \frac{\bar{\mu}_{\zeta} \sigma_{\zeta}^2 + \varphi_{\zeta} \bar{\sigma}_{\zeta}^2 \zeta(t_2)}{\sigma_{\zeta}^2 + \varphi_{\zeta} \bar{\sigma}_{\zeta}^2} \right)^2 \right\},
\]

(2.101)
for $k = 2, \ldots, K - 1,$

$$\pi(\zeta(t_k)|\cdot) \propto |\Delta(t_k)|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2}(\psi(t_k) - \Pi \psi(t_{k-1}))^T \Delta^{-1}(t_k)(\psi(t_k) - \Pi \psi(t_{k-1})) \right\}$$

$$\times \exp \left\{ -\frac{1 + \psi^2}{2\sigma^2_{\zeta}} \left( \zeta(t_k) - \frac{\varphi \zeta(\zeta(t_{k-1}) + \zeta(t_{k+1}))}{1 + \varphi^2_{\zeta}} \right)^2 \right\},$$

and for $k = K,$

$$\pi(\zeta(t_K)|\cdot) \propto |\Delta(t_K)|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2}(\psi(t_K) - \Pi \psi(t_{K-1}))^T \Delta^{-1}(t_K)(\psi(t_K) - \Pi \psi(t_{K-1})) \right\}$$

$$\times \exp \left\{ -\frac{1 + \psi^2}{2\sigma^2_{\zeta}} (\zeta(t_K) - \varphi \zeta(t_{K-1}))^2 \right\}.$$

We note that the M-H ratio for updating $\zeta(t_k)$ is given by

$$H^k_\zeta = \frac{|\Delta^{(prop)}(t_k)|^{\frac{1}{2}} \exp \{ -\frac{1}{2}((\psi(t_k) - \Pi \psi(t_{k-1}))^T(\Delta^{(prop)}(t_k))^{-1}(\psi(t_k) - \Pi \psi(t_{k-1}))) \}}{|\Delta(t_k)|^{\frac{1}{2}} \exp \{ -\frac{1}{2}((\psi(t_k) - \Pi \psi(t_{k-1}))^T \Delta^{-1}(t_k)(\psi(t_k) - \Pi \psi(t_{k-1})) \}}.$$

where $\Delta^{(prop)}(t_k) = K(t_k) \Gamma^{(prop)}(t_k) \Gamma^{(prop)}(t_k)^T K(t_k),$ with $K(t_k) = \text{diag}(\kappa_1(t_k), \kappa_2(t_k)),$ and $\Gamma^{(prop)}(t_k) = \begin{pmatrix} 1 & 0 \\ \zeta^{(prop)}(t_k) & 1 \end{pmatrix}.$ The proposal distributions are normals with the following parameter means and variances. For $k = 1,$ we set

$$\text{mean} = \frac{\tilde{\mu}_{\zeta} \sigma^2_{\zeta} + \varphi \zeta \sigma^2_{\zeta}(t_2)}{\sigma^2_{\zeta} + \varphi^2 \sigma^2_{\zeta}}, \quad \text{variance} = \frac{\sigma^2_{\zeta} \sigma^2_{\zeta}}{\sigma^2_{\zeta} + \varphi^2 \sigma^2_{\zeta}};$$

for $k = 2, \ldots, K - 1,$

$$\text{mean} = \frac{\varphi \zeta(\zeta(t_{k-1}) + \zeta(t_{k+1}))}{1 + \varphi^2_{\zeta}}, \quad \text{variance} = \frac{\sigma^2_{\zeta}}{1 + \varphi^2_{\zeta}}.$$
and for $k = K$,\[\text{mean} = \varphi_\zeta(t_{K-1}), \text{variance} = \sigma^2_\zeta. \quad (2.107)\]

Full-conditional distributions of all the hyper-parameters related to $\Delta(t_k)$: It is straightforward to see that

\[
\begin{align*}
\sigma^2_{\kappa_1} | \cdot & \sim IG \left( \frac{K + 2a_{\kappa_1} - 1}{2}, \left[ \frac{1}{2} \sum_{k=1}^{K} \log(\kappa_1(t_k)) - \varphi_{\kappa_1} \log(\kappa_1(t_{K-1})) \right]^2 + 1/b_{\kappa_1} \right)^{-1}, \quad (2.108) \\
\sigma^2_{\kappa_2} | \cdot & \sim IG \left( \frac{K + 2a_{\kappa_2} - 1}{2}, \left[ \frac{1}{2} \sum_{k=1}^{K} (\log(\kappa_2(t_k)) - \varphi_{\kappa_2} \log(\kappa_2(t_{K-1})) \right]^2 + 1/b_{\kappa_2} \right)^{-1}, \quad (2.109) \\
\sigma^2_\zeta | \cdot & \sim IG \left( \frac{K + 2a_\zeta - 1}{2}, \left[ \frac{1}{2} \sum_{k=1}^{K} (\zeta(t_k) - \varphi_\zeta(\zeta(t_{K-1})) \right]^2 + 1/b_\zeta \right)^{-1}. \quad (2.110)
\end{align*}
\]

Moreover, we have $\tilde{\sigma}^2_{\kappa_1} | \cdot \sim IG(\tilde{A}_{\kappa_1}, \tilde{B}_{\kappa_1})$, where

\[
\tilde{A}_{\kappa_1} = \frac{1}{2}[2\tilde{a}_{\kappa_1} + 1], \quad \tilde{B}_{\kappa_1} = \left[ \frac{1}{2} \sum_{k=1}^{K} \left( \log(\kappa_1(t_k)) - \tilde{\mu}_{\kappa_1} \right)^2 + 1/b_{\kappa_1} \right]^{-1}, \quad (2.111)
\]

$\tilde{\sigma}^2_{\kappa_2} | \cdot \sim IG(\tilde{A}_{\kappa_2}, \tilde{B}_{\kappa_2})$, where

\[
\tilde{A}_{\kappa_2} = \frac{1}{2}[2\tilde{a}_{\kappa_2} + 1], \quad \tilde{B}_{\kappa_2} = \left[ \frac{1}{2} \sum_{k=1}^{K} \left( \log(\kappa_2(t_k)) - \tilde{\mu}_{\kappa_2} \right)^2 + 1/b_{\kappa_2} \right]^{-1}, \quad (2.112)
\]

$\tilde{\sigma}^2_\zeta | \cdot \sim IG(\tilde{A}_\zeta, \tilde{B}_\zeta)$, where

\[
\tilde{A}_\zeta = \frac{1}{2}[2\tilde{a}_\zeta + 1], \quad \tilde{B}_\zeta = \left[ \frac{1}{2} \sum_{k=1}^{K} \left( \log(\kappa_1(t_k)) - \tilde{\mu}_\zeta \right)^2 + 1/b_\zeta \right]^{-1}, \quad (2.113)
\]

$\tilde{\mu}_{\kappa_1} | \cdot \sim N(M_{\tilde{\mu}_{\kappa_1}}, V_{\tilde{\mu}_{\kappa_1}})$, with

\[
V_{\tilde{\mu}_{\kappa_1}} = \frac{\tilde{\sigma}^2_{\kappa_1}}{\sigma^2_{\kappa_1} + \sigma^*_{\kappa_1}}, \quad (2.114)
\]

and

\[
M_{\tilde{\mu}_{\kappa_1}} = V_{\tilde{\mu}_{\kappa_1}} \log(\kappa_1(t_1)), \quad (2.115)
\]
\( \tilde{\mu}_{\kappa_2} | \sim N(M_{\tilde{\mu}_{\kappa_2}}, V_{\tilde{\mu}_{\kappa_2}}) \), with
\[
V_{\tilde{\mu}_{\kappa_2}} = \frac{\tilde{\sigma}_\kappa^2 \sigma_\kappa^2}{\sigma_\kappa^2 + \sigma_\kappa^4},
\] (2.166)
and
\[
M_{\tilde{\mu}_{\kappa_2}} = V_{\tilde{\mu}_{\kappa_2}} \log(\kappa_2(t_i)),
\] (2.167)
\( \tilde{\mu}_\zeta | \sim N(M_{\tilde{\mu}_\zeta}, V_{\tilde{\mu}_\zeta}) \), with
\[
V_{\tilde{\mu}_\zeta} = \frac{\tilde{\sigma}_\zeta^2 \sigma_\zeta^2}{\sigma_\zeta^2 + \sigma_\zeta^4},
\] (2.168)
and
\[
M_{\tilde{\mu}_\zeta} = V_{\tilde{\mu}_\zeta} \zeta(t_i).
\] (2.169)

Note that when we fit the models, we set \( \varphi_{\kappa_1} = \varphi_{\kappa_2} = \varphi_\zeta = 1 \), everywhere above. In the case where we use \( \varphi_{\kappa_1}, \varphi_{\kappa_2}, \varphi_\zeta \sim \pi(\varphi) \propto 1 \), the full conditionals become
\[
\varphi_{\kappa_1} | \sim N \left( \frac{\sum_{k=2}^K \log(\kappa_1(t_k)) \log(\kappa_1(t_{k-1}))}{\sum_{k=2}^K (\log(\kappa_1(t_{k-1})))^2}, \frac{\sigma_{\kappa_1}^2}{\sum_{k=2}^K (\log(\kappa_1(t_{k-1})))^2} \right),
\] (2.170)
\[
\varphi_{\kappa_2} | \sim N \left( \frac{\sum_{k=2}^K \log(\kappa_2(t_k)) \log(\kappa_2(t_{k-1}))}{\sum_{k=2}^K (\log(\kappa_2(t_{k-1})))^2}, \frac{\sigma_{\kappa_2}^2}{\sum_{k=2}^K (\log(\kappa_2(t_{k-1})))^2} \right),
\] (2.171)
and
\[
\varphi_\zeta | \sim N \left( \frac{\sum_{k=2}^K \zeta(t_k) \zeta(t_{k-1})}{\sum_{k=2}^K (\zeta(t_{k-1}))^2}, \frac{\sigma_\zeta^2}{\sum_{k=2}^K (\zeta(t_{k-1}))^2} \right).
\] (2.172)

2.A.11 Updating the full-conditional distribution of \( B(t_k) \)

The full-conditional distribution for the matrix \( B(t_k) \) of the component covariance matrices at time \( t_k \). Recall that \( B(t_k) = P(t_k)F(t_k)F(t_k)^T P(t_k) \), where \( P(t_k) = \text{diag}(\rho_1(t_k), \rho_2(t_k)) \) and \( F(t_k) = \begin{pmatrix} 1 & 0 \\ \varrho(t_k) & 1 \end{pmatrix} \). The full-conditional distributions of the elements of \( B(t_k) \) are
given as follows.

For $k = 1$, the full-conditional distribution of $\rho_1(t_k)$ is given by

$$
\begin{align*}
\pi(\rho_1(t_1) \cdot) & \propto |B(t_1)|^{-a_1m(t_1)} \exp \left\{ - \sum_{j=1}^{m(t_1)} \text{trace}(B(t_1)\Sigma_j^{-1}(t_1)) \right\} \\
& \times \exp \left\{ - \frac{\sigma_{\rho_1}^2 + \varphi_{\rho_1}^2 \tilde{\sigma}_{\rho_1}^2}{2\sigma_{\rho_1}^2 \tilde{\sigma}_{\rho_1}^2} \left( \log(\rho_1(t_1)) - \frac{\tilde{\mu}_{\rho_1} \sigma_{\rho_1}^2 + \varphi_{\rho_1} \tilde{\sigma}_{\rho_1}^2 \log(\rho_1(t_2))}{\sigma_{\rho_1}^2 + \varphi_{\rho_1} \tilde{\sigma}_{\rho_1}^2} \right)^2 \right\},
\end{align*}
$$

(2.123)

for $k = 2, \ldots, K - 1$,

$$
\begin{align*}
\pi(\rho_1(t_k) \cdot) & \propto |B(t_k)|^{-a_1m(t_k)} \exp \left\{ - \sum_{j=1}^{m(t_k)} \text{trace}(B(t_k)\Sigma_j^{-1}(t_k)) \right\} \\
& \times \exp \left\{ - \frac{1 + \varphi_{\rho_1}^2}{2\sigma_{\rho_1}^2} \left( \log(\rho_1(t_k)) - \frac{\varphi_{\rho_1}(\log(\rho_1(t_{k-1})) + \log(\rho_1(t_{k+1})))}{1 + \varphi_{\rho_1}^2} \right)^2 \right\},
\end{align*}
$$

(2.124)

and for $k = K$,

$$
\begin{align*}
\pi(\rho_1(t_K) \cdot) & \propto |B(t_K)|^{-a_1m(t_K)} \exp \left\{ - \sum_{j=1}^{m(t_K)} \text{trace}(B(t_K)\Sigma_j^{-1}(t_K)) \right\} \\
& \times \exp \left\{ - \frac{1}{2\sigma_{\rho_1}^2} \left( \log(\rho_1(t_K)) - \varphi_{\rho_1} \log(\rho_1(t_{K-1})) \right)^2 \right\}.
\end{align*}
$$

(2.125)

We note that the M-H ratio for updating $\rho_1(t_k)$ is given by

$$
H_{\rho_1}^k = \frac{|B^{(prop)}(t_k)|^{-a_1m(t_k)} \exp \left\{ - \sum_{j=1}^{m(t_k)} \text{trace}(B^{(prop)}(t_k)\Sigma_j^{-1}(t_k)) \right\}}{|B(t_k)|^{-a_1m(t_k)} \exp \left\{ - \sum_{j=1}^{m(t_k)} \text{trace}(B(t_k)\Sigma_j^{-1}(t_k)) \right\}},
$$

(2.126)

where $B^{(prop)}(t_k) = P^{(prop)}(t_k)F(t_k)F(t_k)^TP^{(prop)}(t_k)$, $P^{(prop)}(t_k) = \text{diag}(\mu_1^{(prop)}(t_k), \rho_2(t_k))$, and $F(t_k) = \begin{pmatrix} 1 & 0 \\ g(t_k) & 1 \end{pmatrix}$. The proposal distributions are log-normals with the following parameter means and variances. For $k = 1$,

$$
\text{mean} = \frac{\tilde{\mu}_{\rho_1} \sigma_{\rho_1}^2 + \varphi_{\rho_1} \tilde{\sigma}_{\rho_1}^2 \log(\rho_1(t_2))}{\sigma_{\rho_1}^2 + \varphi_{\rho_1} \tilde{\sigma}_{\rho_1}^2}, \quad \text{variance} = \frac{\sigma_{\rho_1}^2 \tilde{\sigma}_{\rho_1}^2}{\sigma_{\rho_1}^2 + \varphi_{\rho_1} \tilde{\sigma}_{\rho_1}^2},
$$

(2.127)
for $k = 2, \ldots, K - 1$,

$$\text{mean} = \frac{\varphi_{\rho_1} \left( \log(\rho_1(t_{k-1})) + \log(\rho_1(t_{k+1})) \right)}{1 + \varphi_{\rho_1}^2}, \quad \text{variance} = \frac{\sigma_{\rho_1}^2}{1 + \varphi_{\rho_1}^2}, \quad (2.128)$$

and for $k = K$,

$$\text{mean} = \varphi_{\rho_1} \log(\rho_1(t_{K-1})), \quad \text{variance} = \sigma_{\rho_1}^2. \quad (2.129)$$

For $k = 1$, the full-conditional distribution of $\rho_2(t_k)$, is given by

$$\pi(\rho_2(t_1) | \cdot) \propto |B(t_1)|^{-a_1m(t_1)} \exp \left\{ -\sum_{j=1}^{m(t_1)} \text{trace}(B(t_1) \Sigma_j^{-1}(t_1)) \right\} \times \exp \left\{ -\frac{\sigma_{\rho_2}^2 + \varphi_{\rho_2}^2 \sigma_{\rho_2}^2}{2 \sigma_{\rho_2}^2 \bar{\sigma}_{\rho_2}^2} \left( \log(\rho_2(t_1)) - \frac{\bar{\mu}_{\rho_2} \sigma_{\rho_2}^2 + \varphi_{\rho_2} \sigma_{\rho_2}^2 \log(\rho_2(t_2))}{\sigma_{\rho_2}^2 + \varphi_{\rho_2} \bar{\sigma}_{\rho_2}^2} \right)^2 \right\}, \quad (2.130)$$

for $k = 2, \ldots, K - 1$,

$$\pi(\rho_2(t_k) | \cdot) \propto |B(t_k)|^{-a_1m(t_k)} \exp \left\{ -\sum_{j=1}^{m(t_k)} \text{trace}(B(t_k) \Sigma_j^{-1}(t_k)) \right\} \times \exp \left\{ -\frac{1 + \varphi_{\rho_2}^2}{2 \sigma_{\rho_2}^2} \left( \log(\rho_2(t_k)) - \frac{\varphi_{\rho_2} (\log(\rho_2(t_{k-1})) + \log(\rho_2(t_{k+1})))}{1 + \varphi_{\rho_2}^2} \right)^2 \right\}, \quad (2.131)$$

and for $k = K$,

$$\pi(\rho_2(t_K) | \cdot) \propto |B(t_K)|^{-a_1m(t_K)} \exp \left\{ -\sum_{j=1}^{m(t_K)} \text{trace}(B(t_K) \Sigma_j^{-1}(t_K)) \right\} \times \exp \left\{ -\frac{1}{2 \sigma_{\rho_2}^2} \left( \log(\rho_2(t_K)) - \varphi_{\rho_2} \log(\rho_2(t_{K-1})) \right)^2 \right\}. \quad (2.132)$$

We note that the M-H ratio for updating $\rho_2(t_k)$ is given by

$$H_{\rho_2}^k = \frac{|B^{(prop)}(t_k)|^{-a_1m(t_k)} \exp \left\{ -\sum_{j=1}^{m(t_k)} \text{trace}(B^{(prop)}(t_k) \Sigma_j^{-1}(t_k)) \right\}}{|B(t_k)|^{-a_1m(t_k)} \exp \left\{ -\sum_{j=1}^{m(t_k)} \text{trace}(B(t_k) \Sigma_j^{-1}(t_k)) \right\}}, \quad (2.133)$$

76
where $B^{(\text{prop})}(t_k) = P^{(\text{prop})}(t_k)F(t_k)F(t_k)^TP^{(\text{prop})}(t_k)$, $P^{(\text{prop})}(t_k) = \text{diag}(\rho_1(t_k), \rho_2(t_k))$, and $F(t_k) = \begin{pmatrix} 1 & 0 \\ g(t_k) & 1 \end{pmatrix}$. The proposal distributions are log-normals with the following parameter means and variances. For $k = 1$, we set

$$
\text{mean} = \frac{\bar{\mu}_2 \sigma_2^2 + \varphi_2 \bar{\sigma}_2^2 \log(\rho_2(t_2))}{\sigma_2^2 + \varphi_2 \bar{\sigma}_2^2}, \quad \text{variance} = \sigma_2^2 \varphi_2 \bar{\sigma}_2^2 + \varphi_2 \bar{\sigma}_2^2,
$$

(2.134)

for $k = 2, \ldots, K - 1$,

$$
\text{mean} = \frac{\varphi_2 (\log(\rho_2(t_{k-1})) + \log(\rho_2(t_{k+1})))}{1 + \varphi_2^2}, \quad \text{variance} = \frac{\sigma_2^2}{1 + \varphi_2^2},
$$

(2.135)

and for $k = K$,

$$
\text{mean} = \varphi_2 \log(\rho_2(t_{K-1})), \quad \text{variance} = \sigma_2^2.
$$

(2.136)

For $k = 1$, the full-conditional distribution of $\varrho(t_k)$, is given by

$$
\pi(\varrho(t_1) | \cdot) \propto |B(t_1)|^{-a_1 m(t_1)} \exp \left\{ - \sum_{j=1}^{m(t_1)} \text{trace}(B(t_1) \Sigma_j^{-1}(t_1)) \right\} \times \exp \left\{ -\frac{\sigma_2^2 + \varphi_2 \bar{\sigma}_2^2}{2\sigma_2^2 \bar{\sigma}_2^2} \left( \varrho(t_1) - \frac{\bar{\mu}_2 \sigma_2^2 + \varphi_2 \bar{\sigma}_2^2 \log(\rho_2(t_2))}{\sigma_2^2 + \varphi_2 \bar{\sigma}_2^2} \right)^2 \right\},
$$

(2.137)

for $k = 2, \ldots, K - 1$,

$$
\pi(\varrho(t_k) | \cdot) \propto |B(t_k)|^{-a_1 m(t_k)} \exp \left\{ - \sum_{j=1}^{m(t_k)} \text{trace}(B(t_k) \Sigma_j^{-1}(t_k)) \right\} \times \exp \left\{ -\frac{1 + \varphi_2^2}{2\sigma_2^2} \left( \varrho(t_k) - \frac{\varphi_2 \varrho(t_{k-1}) + \varrho(t_{k+1})}{1 + \varphi_2^2} \right)^2 \right\},
$$

(2.138)

and for $k = K$,

$$
\pi(\varrho(t_K) | \cdot) \propto |B(t_K)|^{-a_1 m(t_K)} \exp \left\{ - \sum_{j=1}^{m(t_K)} \text{trace}(B(t_K) \Sigma_j^{-1}(t_K)) \right\} \times \exp \left\{ -\frac{1}{2\sigma_2^2} \left( \varrho(t_K) - \varphi_2 \varrho(t_{K-1}) \right)^2 \right\}.
$$

(2.139)
The M-H ratio for updating $\varrho(t_k)$ is given by

$$H^k_{\rho_1} = \frac{|B^{(prop)}(t_k)|^{-\alpha_1 m(t_k)} \exp \left\{ - \sum_{j=1}^{m(t_k)} \text{trace}(B^{(prop)}(t_k)\Sigma_j^{-1}(t_k)) \right\}}{|B(t_k)|^{-\alpha_1 m(t_k)} \exp \left\{ - \sum_{j=1}^{m(t_k)} \text{trace}(B(t_k)\Sigma_j^{-1}(t_k)) \right\}}, \quad (2.140)$$

where $B^{(prop)}(t_k) = P^{(prop)}(t_k)F(t_k)F(t_k)^TP^{(prop)}(t_k)$, $P(t_k) = \text{diag}(\rho_1(t_k), \rho_2(t_k))$, and $F^{(prop)}(t_k) = \begin{pmatrix} 1 & 0 \\ \varrho^{(prop)}(t_k) & 1 \end{pmatrix}$. The proposal distributions are normals with the following parameter means and variances. For $k = 1$, we have

$$\text{mean} = \frac{\mu_{\varrho}_\varrho \sigma^2_{\varrho} + \varrho_{\varrho} \check{\sigma}^2_{\varrho} \xi(t_2)}{\sigma^2_{\varrho} + \varrho^2_{\varrho} \check{\sigma}^2_{\varrho}}, \quad \text{variance} = \frac{\sigma^2_{\varrho} \check{\sigma}^2_{\varrho}}{\sigma^2_{\varrho} + \varrho^2_{\varrho} \check{\sigma}^2_{\varrho}}, \quad (2.141)$$

for $k = 2, \ldots, K-1$,

$$\text{mean} = \frac{\varrho_{\varrho}(\varrho(t_{k-1}) + \varrho(t_{k+1}))}{1 + \varrho^2_{\varrho}}, \quad \text{variance} = \frac{\sigma^2_{\varrho}}{1 + \varrho^2_{\varrho}}, \quad (2.142)$$

for $k = K$,

$$\text{mean} = \varrho_{\varrho}(t_{K-1}), \quad \text{variance} = \sigma^2_{\varrho}. \quad (2.143)$$

Full conditional distributions of all the hyper-parameters related to $B(t_k)$:

It is straightforward to see that

$$\sigma^2_{\rho_1} \mid \sim IG \left( \frac{K + 2a_{\rho_1} - 1}{2}, \left[ \frac{1}{2} \sum_{k=1}^{K} \log(\rho_1(t_k)) - \varrho_{\rho_1} \log(\rho_1(t_{k-1})))^2 + 1/b_{\rho_1} \right]^{-1} \right), \quad (2.144)$$

$$\sigma^2_{\rho_2} \mid \sim IG \left( \frac{K + 2a_{\rho_2} - 1}{2}, \left[ \frac{1}{2} \sum_{k=1}^{K} \log(\rho_2(t_k)) - \varrho_{\rho_2} \log(\rho_2(t_{k-1})))^2 + 1/b_{\rho_2} \right]^{-1} \right), \quad (2.145)$$

$$\sigma^2_{\varrho} \mid \sim IG \left( \frac{K + 2a_{\varrho} - 1}{2}, \left[ \frac{1}{2} \sum_{k=1}^{K} (\varrho(t_k) - \varrho_{\varrho}(t_{k-1}))^2 + 1/b_{\varrho} \right]^{-1} \right), \quad (2.146)$$
\[ \tilde{\sigma}_{\rho_2}^2 \sim IG(\tilde{A}_{\rho_2}, \tilde{B}_{\rho_2}), \]

where

\[ \tilde{A}_{\kappa_2} = \frac{1}{2}[2\tilde{\alpha}_{\rho_2} + 1], \quad \tilde{B}_{\rho_2} = \left[ \frac{1}{2} \left( \log(\rho_2(t_1)) - \tilde{\mu}_{\rho_2} \right)^2 + 1/\tilde{b}_{\rho_2} \right]^{-1}, \tag{2.147} \]

\[ \tilde{\sigma}_\varphi^2 \sim IG(\tilde{A}_\varphi, \tilde{B}_\varphi), \]

where

\[ \tilde{A}_\varphi = \frac{1}{2}[2\tilde{\alpha}_\varphi + 1], \quad \tilde{B}_\varphi = \left[ \frac{1}{2} \left( \rho(t_1) - \tilde{\mu}_\varphi \right)^2 + 1/\tilde{b}_\varphi \right]^{-1}, \tag{2.148} \]

\[ \tilde{\mu}_{\rho_1} \sim N(M_{\tilde{\mu}_{\rho_1}}, V_{\tilde{\mu}_{\rho_1}}), \]

where

\[ V_{\tilde{\mu}_{\rho_1}} = \frac{\tilde{\sigma}_{\rho_1}^2 \sigma_{\rho_1}^*}{\tilde{\sigma}_{\rho_1}^2 + \sigma_{\rho_1}^*} \tag{2.149} \]

and

\[ M_{\tilde{\mu}_{\rho_1}} = V_{\tilde{\mu}_{\rho_1}} \log(\kappa_1(t_1)), \tag{2.150} \]

\[ \tilde{\mu}_{\rho_2} \sim N(M_{\tilde{\mu}_{\rho_2}}, V_{\tilde{\mu}_{\rho_2}}), \]

where

\[ V_{\tilde{\mu}_{\rho_2}} = \frac{\tilde{\sigma}_{\rho_2}^2 \sigma_{\rho_2}^*}{\tilde{\sigma}_{\rho_2}^2 + \sigma_{\rho_2}^*} \tag{2.151} \]

and

\[ M_{\tilde{\mu}_{\rho_2}} = V_{\tilde{\mu}_{\rho_2}} \log(\rho_2(t_1)), \tag{2.152} \]

\[ \tilde{\mu}_\varphi \sim N(M_{\tilde{\mu}_\varphi}, V_{\tilde{\mu}_\varphi}), \]

where

\[ V_{\tilde{\mu}_\varphi} = \frac{\tilde{\sigma}_\varphi^2 \sigma_\varphi^*}{\tilde{\sigma}_\varphi^2 + \sigma_\varphi^*}, \tag{2.153} \]

and

\[ M_{\tilde{\mu}_\varphi} = V_{\tilde{\mu}_\varphi} \varphi(t_1), \tag{2.154} \]

and

\[ \tilde{\sigma}_{\rho_1}^2 \sim IG(\tilde{A}_{\rho_1}, \tilde{B}_{\rho_1}), \]

where

\[ \tilde{A}_{\rho_1} = \frac{1}{2}[2\tilde{\alpha}_{\rho_1} + 1], \quad \tilde{B}_{\rho_1} = \left[ \frac{1}{2} \left( \log(\rho_1(t_1)) - \tilde{\mu}_{\rho_1} \right)^2 + 1/\tilde{b}_{\rho_1} \right]^{-1}. \tag{2.155} \]

Note that when we fit the models, we set \( \varphi_{\rho_1} = \varphi_{\rho_2} = \varphi_{\varphi} = 1 \), everywhere above. In the
case where we use $\varphi_{\rho_1}, \varphi_{\rho_2}, \varphi_{\rho} \sim \pi(\varphi) \propto 1$, the full conditionals become

\[
\begin{align*}
\varphi_{\rho_1} \mid \cdot & \sim N \left( \frac{\sum_{k=2}^{K} \log(\rho_1(t_k)) \log(\rho_1(t_{k-1}))}{\sum_{k=2}^{K} (\log(\rho_1(t_{k-1})))^2}, \frac{\sigma^2_{\rho_1}}{\sum_{k=2}^{K} (\log(\rho_1(t_{k-1})))^2} \right), \\
\varphi_{\rho_2} \mid \cdot & \sim N \left( \frac{\sum_{k=2}^{K} \log(\rho_2(t_k)) \log(\rho_2(t_{k-1}))}{\sum_{k=2}^{K} (\log(\rho_2(t_{k-1})))^2}, \frac{\sigma^2_{\rho_2}}{\sum_{k=2}^{K} (\log(\rho_2(t_{k-1})))^2} \right),
\end{align*}
\]

and

\[
\begin{align*}
\varphi_{\rho} \mid \cdot & \sim N \left( \frac{\sum_{k=2}^{K} \varrho(t_k) \varrho(t_{k-1})}{\sum_{k=2}^{K} (\varrho(t_{k-1}))^2}, \frac{\sigma^2_{\varrho}}{\sum_{k=2}^{K} (\varrho(t_{k-1}))^2} \right).
\end{align*}
\]
2.B Tables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.5000</td>
<td>0.4264</td>
<td>[0.2751, 0.5993]</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.3000</td>
<td>0.2935</td>
<td>[0.1596, 0.4731]</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.3000</td>
<td>0.3419</td>
<td>[0.2509, 0.5113]</td>
</tr>
<tr>
<td>$\Lambda(T)$</td>
<td>32.5249</td>
<td>28.7321</td>
<td>[19.0892, 40.5139]</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>10000.0000</td>
<td>10123.2409</td>
<td>[9929.1027, 10317.133]</td>
</tr>
</tbody>
</table>

Table 2.1: Model 1 (Simulation): Summary statistics for the Hawkes process parameters.

<table>
<thead>
<tr>
<th>Component</th>
<th>Parameter</th>
<th>True Value</th>
<th>Post. Mean</th>
<th>95% Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments</td>
<td>$\lambda$</td>
<td>293.0149</td>
<td>299.8749</td>
<td>[277.4762, 323.7232]</td>
</tr>
<tr>
<td>1 (1)</td>
<td>$p_1$</td>
<td>0.3836</td>
<td>0.3926</td>
<td>[0.3204, 0.4657]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-0.8655</td>
<td>-0.8248</td>
<td>[-0.9953, -0.6612]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>1.4555</td>
<td>1.6508</td>
<td>[1.0782, 2.2332]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.7855</td>
<td>0.55</td>
<td>[0.3859, 0.7311]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>1.2461</td>
<td>1.0933</td>
<td>[0.6722, 1.6194]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>5.1006</td>
<td>5.2897</td>
<td>[3.6902, 7.2183]</td>
</tr>
<tr>
<td>2 (2)</td>
<td>$p_2$</td>
<td>0.6164</td>
<td>0.6074</td>
<td>[0.5343, 0.6796]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>1.0047</td>
<td>1.1686</td>
<td>[0.8111, 1.5377]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>-2.7346</td>
<td>-2.7458</td>
<td>[-3.0381, -2.455]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>4.312</td>
<td>4.2957</td>
<td>[3.4254, 5.3051]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.495</td>
<td>-0.0477</td>
<td>[-0.6661, 0.5709]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>5.1006</td>
<td>5.2897</td>
<td>[3.6902, 7.2183]</td>
</tr>
</tbody>
</table>

Table 2.2: Model 1 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=1 ($t_1 = 2.6522$). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis).

<table>
<thead>
<tr>
<th>Component</th>
<th>Parameter</th>
<th>True Value</th>
<th>Post. Mean</th>
<th>95% Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments</td>
<td>$\lambda$</td>
<td>347.7975</td>
<td>351.0332</td>
<td>[342.7968, 359.5376]</td>
</tr>
<tr>
<td>1 (1)</td>
<td>$p_1$</td>
<td>0.3504</td>
<td>0.4075</td>
<td>[0.3156, 0.4923]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-0.9903</td>
<td>-0.9891</td>
<td>[-1.8668, -0.6515]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>4.8679</td>
<td>4.6912</td>
<td>[4.242, 5.7951]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.9768</td>
<td>1.1842</td>
<td>[0.733, 2.4557]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.8317</td>
<td>-1.066</td>
<td>[-2.4344, -0.5664]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>1.5961</td>
<td>1.8637</td>
<td>[1.1971, 3.2367]</td>
</tr>
<tr>
<td>2 (2)</td>
<td>$p_2$</td>
<td>0.0976</td>
<td>0.0952</td>
<td>[0.0529, 0.2175]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-3.9585</td>
<td>-3.3831</td>
<td>[-3.9962, -0.1839]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>7.6673</td>
<td>7.3878</td>
<td>[3.2018, 8.0908]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.192</td>
<td>0.3153</td>
<td>[0.1353, 0.9948]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>0.1188</td>
<td>0.1201</td>
<td>[-0.2895, 0.3573]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.1697</td>
<td>0.3702</td>
<td>[0.1546, 1.2347]</td>
</tr>
</tbody>
</table>

Continued on next page
Table 2.3: Model 1 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=13 ($t_{13} = 25.3369$). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis).

<table>
<thead>
<tr>
<th>Component Assignments</th>
<th>Parameter</th>
<th>True Value</th>
<th>Post. Mean</th>
<th>95% Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (2)</td>
<td>$\lambda$</td>
<td>446.7691</td>
<td>453.643</td>
<td>[436.1862, 470.6695]</td>
</tr>
<tr>
<td></td>
<td>$p_1$</td>
<td>0.396</td>
<td>0.5145</td>
<td>[0.4508, 0.578]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>0.5836</td>
<td>-0.4542</td>
<td>[-0.5973, -0.3069]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>-21.6229</td>
<td>-17.6169</td>
<td>[-17.8536, -17.3829]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.9958</td>
<td>1.0598</td>
<td>[0.8695, 1.2766]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-1.6456</td>
<td>-1.37</td>
<td>[-1.6871, -1.104]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>2.8347</td>
<td>2.7827</td>
<td>[2.2819, 3.3416]</td>
</tr>
<tr>
<td>2 (1)</td>
<td>$p_2$</td>
<td>0.604</td>
<td>0.4855</td>
<td>[0.422, 0.5492]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-0.4589</td>
<td>0.3829</td>
<td>[0.1824, 0.5853]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>-17.7878</td>
<td>-21.1928</td>
<td>[-21.5636, -20.8113]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>1.1609</td>
<td>1.2965</td>
<td>[1.0103, 1.62]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-1.4181</td>
<td>-2.1858</td>
<td>[-2.7795, -1.6767]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>2.7382</td>
<td>3.9905</td>
<td>[3.0414, 5.0782]</td>
</tr>
</tbody>
</table>

Table 2.4: Model 1 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=27 ($t_{27} = 47.6197$). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis). Note the index change from 1 to 2, and 2 to 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.5000</td>
<td>0.4799</td>
<td>[0.2351, 0.7283]</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.3000</td>
<td>0.2809</td>
<td>[0.0092, 0.8891]</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.3000</td>
<td>0.2829</td>
<td>[0.0017, 1.1985]</td>
</tr>
<tr>
<td>$\Lambda(T)$</td>
<td>32.5249</td>
<td>28.9516</td>
<td>[19.2511, 40.1639]</td>
</tr>
</tbody>
</table>

Table 2.5: Model 2 (Simulation): Summary statistics for the Hawkes process parameters.
Table 2.6 – Continued from previous page

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>µ₂</td>
<td></td>
<td>-5.4312</td>
<td>-5.3991</td>
<td>[-5.5456, -5.2473]</td>
</tr>
<tr>
<td>σ₁₁</td>
<td></td>
<td>12.4303</td>
<td>13.2039</td>
<td>[10.579, 16.2911]</td>
</tr>
<tr>
<td>σ₁₂</td>
<td></td>
<td>-2.0497</td>
<td>-1.6111</td>
<td>[-2.2084, -1.0983]</td>
</tr>
<tr>
<td>σ₂₂</td>
<td></td>
<td>1.0802</td>
<td>0.8773</td>
<td>[0.6917, 1.0833]</td>
</tr>
<tr>
<td></td>
<td>2 (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>µ₁</td>
<td></td>
<td>-8.3821</td>
<td>-8.5295</td>
<td>[-8.7279, -8.3256]</td>
</tr>
<tr>
<td>σ₁₁</td>
<td></td>
<td>0.0936</td>
<td>0.1439</td>
<td>[0.0822, 0.1287]</td>
</tr>
<tr>
<td>σ₁₂</td>
<td></td>
<td>0.1899</td>
<td>0.1914</td>
<td>[0.1261, 0.2683]</td>
</tr>
<tr>
<td>σ₂₂</td>
<td></td>
<td>1.7312</td>
<td>1.688</td>
<td>[1.3307, 2.1024]</td>
</tr>
</tbody>
</table>

Table 2.6: Model 2 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=1 ($t_1 = 2.6522$). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis).

<table>
<thead>
<tr>
<th>Component</th>
<th>Parameter</th>
<th>True Value</th>
<th>Post. Mean</th>
<th>95% Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments</td>
<td>λ</td>
<td>319.568</td>
<td>325.8479</td>
<td>[290.8989, 361.883]</td>
</tr>
<tr>
<td>1 (3)</td>
<td>$p_1$</td>
<td>0.4257</td>
<td>0.2159</td>
<td>[0.1711, 0.2643]</td>
</tr>
<tr>
<td>µ₁</td>
<td>-4.3595</td>
<td>0.1444</td>
<td>[-0.0334, 0.3226]</td>
<td></td>
</tr>
<tr>
<td>σ₁₁</td>
<td>0.0551</td>
<td>0.5057</td>
<td>[0.3445, 0.7044]</td>
<td></td>
</tr>
<tr>
<td>σ₁₂</td>
<td>1.5485</td>
<td>3.2239</td>
<td>[2.244, 4.5571]</td>
<td></td>
</tr>
<tr>
<td>σ₂₂</td>
<td>54.63</td>
<td>28.644</td>
<td>[20.5611, 39.1888]</td>
<td></td>
</tr>
<tr>
<td>2 (2)</td>
<td>$p_2$</td>
<td>0.3231</td>
<td>0.3307</td>
<td>[0.2778, 0.3854]</td>
</tr>
<tr>
<td>µ₁</td>
<td>1.5423</td>
<td>1.5475</td>
<td>[1.4879, 1.6012]</td>
<td></td>
</tr>
<tr>
<td>µ₂</td>
<td>-15.7049</td>
<td>-15.5156</td>
<td>[-15.9901, -15.0032]</td>
<td></td>
</tr>
<tr>
<td>σ₁₁</td>
<td>0.0217</td>
<td>0.0807</td>
<td>[0.0366, 0.1356]</td>
<td></td>
</tr>
<tr>
<td>σ₁₂</td>
<td>-0.1908</td>
<td>-0.1897</td>
<td>[-0.3822, -0.0546]</td>
<td></td>
</tr>
<tr>
<td>σ₂₂</td>
<td>6.6822</td>
<td>5.5695</td>
<td>[3.9005, 7.4895]</td>
<td></td>
</tr>
<tr>
<td>3 (1)</td>
<td>$p_3$</td>
<td>0.2423</td>
<td>0.4534</td>
<td>[0.3993, 0.5087]</td>
</tr>
<tr>
<td>µ₁</td>
<td>0.1132</td>
<td>-4.3618</td>
<td>[-4.4085, -4.3175]</td>
<td></td>
</tr>
<tr>
<td>σ₁₁</td>
<td>0.4325</td>
<td>0.0858</td>
<td>[0.052, 0.1189]</td>
<td></td>
</tr>
<tr>
<td>σ₁₂</td>
<td>3.2524</td>
<td>1.266</td>
<td>[0.9344, 1.6764]</td>
<td></td>
</tr>
<tr>
<td>σ₂₂</td>
<td>28.344</td>
<td>46.8589</td>
<td>[37.2216, 58.5766]</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.7: Model 2 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=13 ($t_{13} = 25.3369$). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis). Note the index change from 1 to 3, and 3 to 1.

<table>
<thead>
<tr>
<th>Component</th>
<th>Parameter</th>
<th>True Value</th>
<th>Post. Mean</th>
<th>95% Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments</td>
<td>λ</td>
<td>336.2249</td>
<td>349.079</td>
<td>[313.6038, 386.31]</td>
</tr>
</tbody>
</table>
Table 2.8: Model 2 (Simulation): Summary of the true and estimated values of the mixture parameters at time index=27 ($t_{27} = 47.6197$). The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis). Note the index change from 1 to 2, and 2 to 1.

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (2)</td>
<td>$p_1$</td>
<td>0.402</td>
<td>0.5643</td>
<td>[0.5119, 0.6159]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>3.4264</td>
<td>5.7078</td>
<td>[5.6602, 5.7559]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>-21.5673</td>
<td>-18.6867</td>
<td>[-19.7228, -17.685]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.2954</td>
<td>0.1181</td>
<td>[0.0797, 0.1725]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>6.4477</td>
<td>0.5827</td>
<td>[0.2376, 0.959]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>149.6502</td>
<td>54.9828</td>
<td>[44.6044, 66.4041]</td>
</tr>
<tr>
<td>2 (1)</td>
<td>$p_2$</td>
<td>0.598</td>
<td>0.4357</td>
<td>[0.3841, 0.4881]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>5.6824</td>
<td>3.408</td>
<td>[3.3104, 3.5022]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>-18.6459</td>
<td>-21.9892</td>
<td>[-23.995, -20.0162]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0699</td>
<td>0.3651</td>
<td>[0.2807, 0.4769]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>0.753</td>
<td>6.7478</td>
<td>[5.3045, 8.5654]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>57.5566</td>
<td>157.9186</td>
<td>[125.9949, 197.1378]</td>
</tr>
</tbody>
</table>

Table 2.9: Model 1 (Earthquakes application): Posterior estimates for the Hawkes process parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.0085</td>
<td>[2e-04, 0.0309]</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.8916</td>
<td>[0.7025, 0.9953]</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.4759</td>
<td>[0.4175, 0.5379]</td>
</tr>
<tr>
<td>$\Lambda(T)$</td>
<td>40.7451</td>
<td>[32.334, 46.6464]</td>
</tr>
</tbody>
</table>

Table 2.8 – Continued from previous page

Continued on next page
Table 2.10: Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=1.

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>118.149</td>
<td>λ</td>
<td>264.2447</td>
<td>[259.776, 269.2943]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>p_1</td>
<td>0.1878</td>
<td>[0.0285, 0.2408]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ_1</td>
<td>-115.4824</td>
<td>[-115.5609, -114.544]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ_2</td>
<td>32.6476</td>
<td>[32.512, 35.8722]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ_{11}</td>
<td>0.0349</td>
<td>[0.0223, 0.0543]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ_{12}</td>
<td>-0.0263</td>
<td>[-0.0404, -0.0155]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ_{22}</td>
<td>0.0208</td>
<td>[0.0092, 0.0315]</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>p_2</td>
<td>0.0212</td>
<td>[0.004, 0.1715]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ_1</td>
<td>-114.7259</td>
<td>[-115.5015, -114.5359]</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>$p_3$</td>
<td>0.1225</td>
<td></td>
<td>[0.011, 0.2951]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-116.1577</td>
<td></td>
<td>[-116.3606, -115.8533]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.6612</td>
<td></td>
<td>[32.0149, 33.6885]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.3308</td>
<td></td>
<td>[0.0034, 0.947]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.4627</td>
<td></td>
<td>[-1.3672, -2e-04]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.7786</td>
<td></td>
<td>[5e-04, 2.3154]</td>
</tr>
<tr>
<td>4</td>
<td>$p_4$</td>
<td>0.1579</td>
<td></td>
<td>[0.0202, 0.2953]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-116.5505</td>
<td></td>
<td>[-117.683, -115.9518]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.2268</td>
<td></td>
<td>[32.0578, 34.0992]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.716</td>
<td></td>
<td>[0.0063, 2.0833]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.7053</td>
<td></td>
<td>[-1.458, -0.0018]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>1.0481</td>
<td></td>
<td>[0.001, 2.4074]</td>
</tr>
<tr>
<td>5</td>
<td>$p_5$</td>
<td>0.0539</td>
<td></td>
<td>[0.0255, 0.088]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.3738</td>
<td></td>
<td>[-118.1621, -117.123]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.8738</td>
<td></td>
<td>[32.5791, 34.1546]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.252</td>
<td></td>
<td>[0.0282, 1.8081]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.088</td>
<td></td>
<td>[-0.6352, -0.0027]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0382</td>
<td></td>
<td>[0.003, 0.2345]</td>
</tr>
<tr>
<td>6</td>
<td>$p_6$</td>
<td>0.3002</td>
<td></td>
<td>[0.2482, 0.3588]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.8598</td>
<td></td>
<td>[-117.8608, -117.855]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>36.3874</td>
<td></td>
<td>[36.3887, 36.3939]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0035</td>
<td></td>
<td>[1e-04, 3e-04]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0012</td>
<td></td>
<td>[-1e-04, 0]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>6e-04</td>
<td></td>
<td>[1e-04, 2e-04]</td>
</tr>
<tr>
<td>7</td>
<td>$p_7$</td>
<td>0.1566</td>
<td></td>
<td>[0.1114, 0.2072]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-120.006</td>
<td></td>
<td>[-120.4729, -119.5449]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>35.0561</td>
<td></td>
<td>[34.5958, 35.5398]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>1.7811</td>
<td></td>
<td>[1.0874, 2.9329]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-1.3903</td>
<td></td>
<td>[-2.2753, -0.7711]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>1.6005</td>
<td></td>
<td>[0.8876, 2.6123]</td>
</tr>
</tbody>
</table>

Table 2.11: Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=26.
<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$p_2$</td>
<td>0.185</td>
<td>0.0031</td>
<td>[ 0.1301 , 0.2706 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.5027</td>
<td>-115.8114 , -115.4137</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.5484</td>
<td>32.4113 , 32.8552</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.053</td>
<td>0.0209 , 0.2682</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0489</td>
<td>-0.2617 , -0.0102</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.1541</td>
<td>0.0864 , 0.35</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$p_3$</td>
<td>0.0319</td>
<td></td>
<td>[ 0.0105 , 0.0644 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-116.7461</td>
<td>-117.2086 , -116.5555</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.2269</td>
<td>32.1053 , 32.2778</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0933</td>
<td>5e-04 , 0.5543</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>0.0128</td>
<td>-0.0023 , 0.0787</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0099</td>
<td>3e-04 , 0.027</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$p_4$</td>
<td>0.1406</td>
<td></td>
<td>[ 0.0551 , 0.2713 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.1106</td>
<td>-117.2193 , -116.8053</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.869</td>
<td>33.4636 , 34.1643</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.4195</td>
<td>2e-04 , 1.1606</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.1241</td>
<td>-0.6451 , 0.1638</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.4936</td>
<td>1e-04 , 1.9337</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>$p_5$</td>
<td>0.1357</td>
<td></td>
<td>[ 0.052 , 0.26 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.2519</td>
<td>-117.6721 , -117.1684</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>34.0509</td>
<td>33.6425 , 34.4278</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.3081</td>
<td>2e-04 , 0.982</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.1055</td>
<td>-0.5761 , 0.3164</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.5753</td>
<td>1e-04 , 1.9325</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>$p_6$</td>
<td>0.1666</td>
<td></td>
<td>[ 0.1114 , 0.2347 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.8267</td>
<td>-117.8616 , -117.7469</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>36.0506</td>
<td>36.0243 , 36.0794</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0107</td>
<td>6e-04 , 0.0363</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.004</td>
<td>-0.0152 , -3e-04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0057</td>
<td>3e-04 , 0.0192</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>$p_7$</td>
<td>0.0783</td>
<td></td>
<td>[ 0.0452 , 0.1179 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-120.8949</td>
<td>-121.143 , -120.3905</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>36.1041</td>
<td>35.8223 , 36.354</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.2798</td>
<td>0.0714 , 1.0818</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.1826</td>
<td>-0.5752 , -0.0498</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.2625</td>
<td>0.1221 , 0.5424</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.12: Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=27.
Table 2.13 – Continued from previous page

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$p_2$</td>
<td>0.1451</td>
<td>[0.071, 0.238]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.2364</td>
<td>[-115.3052, -115.1616]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.1793</td>
<td>[32.1237, 32.2334]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0093</td>
<td>[0.0054, 0.0143]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.005</td>
<td>[-0.0107, 0.004]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0049</td>
<td>[0.0016, 0.0091]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$p_3$</td>
<td>0.2324</td>
<td>[0.1484, 0.295]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.3263</td>
<td>[-115.3343, -115.318]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.2545</td>
<td>[32.2396, 32.2666]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0022</td>
<td>[0.0013, 0.0029]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0014</td>
<td>[-0.002, -0.001]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.004</td>
<td>[0.0028, 0.0058]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$p_4$</td>
<td>0.0922</td>
<td>[0.016, 0.1347]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.6127</td>
<td>[-115.6407, -115.5482]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.4704</td>
<td>[32.3796, 32.5055]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0139</td>
<td>[0.0041, 0.0544]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.012</td>
<td>[-0.0484, -0.0036]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0127</td>
<td>[0.0043, 0.052]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>$p_5$</td>
<td>0.0339</td>
<td>[0.0018, 0.1165]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.6351</td>
<td>[-115.7342, -115.53]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.6309</td>
<td>[32.4118, 33.0655]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0236</td>
<td>[8e-04, 0.0669]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.021</td>
<td>[-0.0641, 0.007]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0284</td>
<td>[0.0023, 0.0826]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>$p_6$</td>
<td>0.0226</td>
<td>[0.0015, 0.2187]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.6381</td>
<td>[-115.7798, -115.5381]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.8161</td>
<td>[32.4648, 33.0979]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0149</td>
<td>[7e-04, 0.0643]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0123</td>
<td>[-0.0627, 0.0077]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0257</td>
<td>[9e-04, 0.0825]</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>$p_7$</td>
<td>0.1912</td>
<td>[0.0041, 0.2444]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.7621</td>
<td>[-115.8069, -115.6482]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.643</td>
<td>[32.5598, 32.8901]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0079</td>
<td>[6e-04, 0.0676]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0075</td>
<td>[-0.0619, -2e-04]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0087</td>
<td>[8e-04, 0.0606]</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>$p_8$</td>
<td>0.0835</td>
<td>[0.0051, 0.2022]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.8298</td>
<td>[-115.8583, -115.7246]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.698</td>
<td>[32.6292, 32.9297]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0106</td>
<td>[7e-04, 0.1014]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0086</td>
<td>[-0.0927, -3e-04]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0082</td>
<td>[4e-04, 0.0898]</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>$p_9$</td>
<td>0.0359</td>
<td>[0.0042, 0.1175]</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.13 – Continued from previous page

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.8827</td>
<td>$[-116.0432, -115.781]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.7238</td>
<td>$[32.6738, 33.065]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0291</td>
<td>$[9e-04, 0.2398]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0201</td>
<td>$[-0.1976, 0.0012]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0185</td>
<td>$[4e-04, 0.1619]$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>$p_{10}$</td>
<td>0.0085</td>
<td>$[0.0014, 0.0381]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.9629</td>
<td>$[-116.2142, -115.8227]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.0455</td>
<td>$[32.6781, 33.3778]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0686</td>
<td>$[7e-04, 0.4585]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.065</td>
<td>$[-0.4523, 0.0011]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0665</td>
<td>$[5e-04, 0.5138]$</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>$p_{11}$</td>
<td>0.0051</td>
<td>$[6e-04, 0.0156]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-116.2037</td>
<td>$[-116.5001, -115.9975]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.3607</td>
<td>$[32.1378, 33.8876]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0268</td>
<td>$[9e-04, 0.1898]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0273</td>
<td>$[-0.2104, 7e-04]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0361</td>
<td>$[5e-04, 0.2352]$</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>$p_{12}$</td>
<td>0.0047</td>
<td>$[7e-04, 0.0119]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-116.3064</td>
<td>$[-116.6562, -116.0855]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.295</td>
<td>$[32.045, 33.872]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0192</td>
<td>$[8e-04, 0.0836]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0183</td>
<td>$[-0.0672, 0.0023]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0263</td>
<td>$[6e-04, 0.108]$</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>$p_{13}$</td>
<td>0.0032</td>
<td>$[9e-04, 0.0078]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-116.4512</td>
<td>$[-116.667, -116.295]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.4108</td>
<td>$[33.045, 33.872]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0187</td>
<td>$[8e-04, 0.0828]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.014</td>
<td>$[-0.0507, 1e-04]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0159</td>
<td>$[6e-04, 0.0609]$</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>$p_{14}$</td>
<td>0.0031</td>
<td>$[6e-04, 0.0108]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.7534</td>
<td>$[-118.0271, -117.6211]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.8619</td>
<td>$[33.6014, 35.1195]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.1194</td>
<td>$[6e-04, 0.8153]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.1466</td>
<td>$[-1.0382, 4e-04]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.2701</td>
<td>$[4e-04, 1.9546]$</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>$p_{15}$</td>
<td>0.0083</td>
<td>$[8e-04, 0.0275]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.9315</td>
<td>$[-118.3747, -117.7134]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>35.2984</td>
<td>$[33.6356, 36.2802]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.6755</td>
<td>$[2e-04, 4.2202]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.4068</td>
<td>$[-2.6089, 0]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.4223</td>
<td>$[2e-04, 2.5617]$</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>$p_{16}$</td>
<td>0.0061</td>
<td>$[0.001, 0.0207]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.9172</td>
<td>$[-118.5029, -117.7948]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>36.0106</td>
<td>$[34.114, 36.4053]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.1783</td>
<td>$[2e-04, 3.246]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.106</td>
<td>$[-1.9519, 8e-04]$</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
Table 2.13 – Continued from previous page

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>$\sigma_{22}$</td>
<td>0.1107</td>
<td>[2e-04, 2.0103]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>$\lambda$</td>
<td>415.4832</td>
<td>[406.11, 424.4162]</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.13: Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=28.

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$p_1$</td>
<td>0.5763</td>
<td>[0.533, 0.6203]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.2844</td>
<td>[-115.3011, -115.2678]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.2193</td>
<td>[32.2065, 32.2324]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.018</td>
<td>[0.0153, 0.0211]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.013</td>
<td>[-0.0154, -0.0109]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0113</td>
<td>[0.0095, 0.0134]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$p_2$</td>
<td>0.2933</td>
<td>[0.1528, 0.3464]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.8034</td>
<td>[-115.8156, -115.7917]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.6633</td>
<td>[32.6537, 32.6713]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0045</td>
<td>[0.0031, 0.0059]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0025</td>
<td>[-0.0034, -0.0018]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.002</td>
<td>[0.0015, 0.0026]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>$p_3$</td>
<td>0.0337</td>
<td>[0.0082, 0.1684]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-116.01</td>
<td>[-116.153, -115.8155]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.7751</td>
<td>[32.6746, 33.0267]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0131</td>
<td>[0.0019, 0.0055]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0084</td>
<td>[-0.0536, 0.0021]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0179</td>
<td>[0.0012, 0.1122]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$p_4$</td>
<td>0.0096</td>
<td>[0.0027, 0.0205]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.5998</td>
<td>[-117.7211, -117.4679]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.2198</td>
<td>[32.1235, 32.3202]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0337</td>
<td>[0.0013, 0.0549]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0187</td>
<td>[-0.0297, 2e-04]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0205</td>
<td>[6e-04, 0.0214]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$p_5$</td>
<td>0.0764</td>
<td>[0.0514, 0.1036]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.7498</td>
<td>[-118.2788, -117.2464]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>34.4302</td>
<td>[34.0457, 34.9117]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>2.425</td>
<td>[1.4423, 4.0218]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-1.0432</td>
<td>[-1.9961, -0.423]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>1.3899</td>
<td>[0.8585, 2.1808]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>$p_6$</td>
<td>0.0108</td>
<td>[0.0032, 0.0219]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-120.3592</td>
<td>[-120.5255, -120.1836]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.8079</td>
<td>[33.7214, 33.8925]</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
Table 2.14 - Continued from previous page

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0211</td>
<td>[0.0043, 0.081]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0081</td>
<td>[-0.0289, -0.0013]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0045</td>
<td>[0.0011, 0.0162]</td>
</tr>
</tbody>
</table>

Table 2.14: Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=29.

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p_1$</td>
<td>0.184</td>
<td>[0.1277, 0.2623]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-115.3936</td>
<td>[-115.4639, -115.3391]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>32.6263</td>
<td>[32.5087, 32.7669]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0406</td>
<td>[0.0262, 0.0665]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0849</td>
<td>[-0.1403, -0.0549]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.1927</td>
<td>[0.1294, 0.3139]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_1$</td>
<td>0.188</td>
<td>[0.1277, 0.2623]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-116.3273</td>
<td>[-116.4907, -116.1757]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.6516</td>
<td>[33.3723, 33.9135]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.1213</td>
<td>[0.0536, 0.2184]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.108</td>
<td>[-0.2185, -0.0313]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.4571</td>
<td>[0.2325, 0.7284]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_1$</td>
<td>0.0555</td>
<td>[0.0071, 0.1346]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.3997</td>
<td>[-118.0343, -116.9014]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.2457</td>
<td>[32.0099, 34.1536]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.2467</td>
<td>[0.0149, 0.7911]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>0.0017</td>
<td>[-0.2913, 0.5044]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0824</td>
<td>[0.0105, 0.4844]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_1$</td>
<td>0.0466</td>
<td>[0.0062, 0.1435]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-117.7621</td>
<td>[-118.4638, -117.0718]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>33.1445</td>
<td>[32.1447, 34.2455]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.2346</td>
<td>[0.0154, 0.7949]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0098</td>
<td>[-0.429, 0.5447]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.156</td>
<td>[0.0113, 0.6688]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_1$</td>
<td>0.0821</td>
<td>[0.0447, 0.12]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-118.8131</td>
<td>[-118.3648, -117.604]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>35.5535</td>
<td>[33.6774, 36.06]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0694</td>
<td>[0.0235, 0.204]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0487</td>
<td>[-0.3233, 0.0038]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0956</td>
<td>[0.0093, 0.6198]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_1$</td>
<td>0.0803</td>
<td>[0.0348, 0.1377]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>-118.3056</td>
<td>[-118.6253, -117.6615]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>34.4473</td>
<td>[33.7791, 36.041]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.1459</td>
<td>[0.0268, 0.5987]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.1497</td>
<td>[-0.4098, 0.0029]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.2964</td>
<td>[0.012, 1.009]</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td></td>
<td>$p_1$</td>
<td>0.174</td>
<td>[0.028, 0.2595]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-118.7095</td>
<td>[-119.381, -118.229]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>35.152</td>
<td>[34.182, 35.7724]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>1.9502</td>
<td>[0.121, 3.3773]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>0.201</td>
<td>[-0.2502, 0.7373]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.7766</td>
<td>[0.1162, 1.4226]</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>$p_1$</td>
<td>0.0577</td>
<td>[0.02, 0.1611]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-120.2354</td>
<td>[-120.5729, -119.1327]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>36.0822</td>
<td>[34.1971, 36.4876]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.3484</td>
<td>[0.05, 3.2035]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0598</td>
<td>[-0.2132, 0.3156]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.1162</td>
<td>[0.0368, 0.434]</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>$p_1$</td>
<td>0.0307</td>
<td>[0.0097, 0.0688]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-121.0277</td>
<td>[-121.3331, -120.3094]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>35.8059</td>
<td>[35.573, 36.4188]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.067</td>
<td>[0.0143, 0.2458]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0268</td>
<td>[-0.1429, 0.0662]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.04</td>
<td>[0.0091, 0.1538]</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>$p_1$</td>
<td>0.1011</td>
<td>[0.07, 0.1392]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-121.3319</td>
<td>[-121.3961, -121.2609]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>36.6901</td>
<td>[36.6405, 36.7354]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0355</td>
<td>[0.0212, 0.0601]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0166</td>
<td>[-0.0299, -0.0078]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0189</td>
<td>[0.0115, 0.0302]</td>
</tr>
</tbody>
</table>

Table 2.15: Model 1 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=44.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.2097</td>
<td>[0.1066, 0.3045]</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.2244</td>
<td>[0.0148, 0.7581]</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.2319</td>
<td>[0.007, 1.0883]</td>
</tr>
<tr>
<td>$\Lambda(T)$</td>
<td>47.8706</td>
<td>[35.2123, 62.1003]</td>
</tr>
</tbody>
</table>

Table 2.16: Model 2 (Earthquakes application): Posterior estimates of the model parameters for the Hawkes process.

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0237</td>
<td></td>
<td>$\lambda$</td>
<td>211.3131</td>
<td>[183.3594, 241.4088]</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>$p_1$</td>
<td>0.0749</td>
<td>[0.0328, 0.1155]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-115.2357</td>
<td>[-115.5681, -115.0333]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>32.3723</td>
<td>[32.1727, 32.9846]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0699</td>
<td>[2e-04, 0.4879]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0575</td>
<td>[-0.4302, 0]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0495</td>
<td>[4e-04, 0.3758]</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td>( p_2 )</td>
<td>0.0826</td>
<td>[0.044, 0.1192]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_1 )</td>
<td>-115.5813</td>
<td>[-115.8929, -115.5417]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_2 )</td>
<td>32.9684</td>
<td>[32.7964, 33.0084]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{11} )</td>
<td>0.0477</td>
<td>[1e-04, 0.5804]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{12} )</td>
<td>-0.0415</td>
<td>[-0.5063, 1e-04]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{22} )</td>
<td>0.0371</td>
<td>[2e-04, 0.446]</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>( p_3 )</td>
<td>0.2504</td>
<td>[0.1789, 0.3491]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_1 )</td>
<td>-116.2825</td>
<td>[-116.3026, -116.2626]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_2 )</td>
<td>34.6359</td>
<td>[34.5964, 34.6766]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{11} )</td>
<td>0.0062</td>
<td>[0.0039, 0.0086]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{12} )</td>
<td>-0.0106</td>
<td>[-0.0153, -0.0066]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{22} )</td>
<td>0.0249</td>
<td>[0.017, 0.0349]</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>( p_4 )</td>
<td>0.1311</td>
<td>[0.0933, 0.1744]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_1 )</td>
<td>-116.7182</td>
<td>[-116.9212, -116.5217]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_2 )</td>
<td>34.2655</td>
<td>[34.1825, 34.3422]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{11} )</td>
<td>0.1651</td>
<td>[0.0728, 0.3047]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{12} )</td>
<td>0.0604</td>
<td>[0.0225, 0.1226]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{22} )</td>
<td>0.0317</td>
<td>[0.012, 0.0585]</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>( p_5 )</td>
<td>0.1781</td>
<td>[0.0451, 0.2733]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_1 )</td>
<td>-117.7251</td>
<td>[-118.0514, -117.2254]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_2 )</td>
<td>34.4963</td>
<td>[33.6241, 35.7334]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{11} )</td>
<td>2.0673</td>
<td>[3e-04, 3.7947]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{12} )</td>
<td>-0.7579</td>
<td>[-1.6735, 0]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{22} )</td>
<td>1.3609</td>
<td>[0.0016, 2.5616]</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>( p_6 )</td>
<td>0.0859</td>
<td>[0.0263, 0.2232]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_1 )</td>
<td>-117.6759</td>
<td>[-118.1231, -117.5178]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_2 )</td>
<td>35.6231</td>
<td>[34.2051, 36.0889]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{11} )</td>
<td>0.296</td>
<td>[2e-04, 2.9718]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{12} )</td>
<td>-0.1056</td>
<td>[-1.1822, 8e-04]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{22} )</td>
<td>0.1846</td>
<td>[3e-04, 2.0087]</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>( p_7 )</td>
<td>0.1089</td>
<td>[0.0411, 0.2126]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_1 )</td>
<td>-117.7069</td>
<td>[-118.2452, -117.597]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_2 )</td>
<td>35.8999</td>
<td>[34.3443, 36.0937]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{11} )</td>
<td>0.2652</td>
<td>[2e-04, 2.9093]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{12} )</td>
<td>-0.0929</td>
<td>[-1.144, 0.0048]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{22} )</td>
<td>0.1769</td>
<td>[2e-04, 2.0748]</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>( p_8 )</td>
<td>0.0881</td>
<td>[0.0166, 0.1461]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_1 )</td>
<td>-117.9725</td>
<td>[-119.0053, -117.7172]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_2 )</td>
<td>36.1436</td>
<td>[35.7182, 37.0099]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{11} )</td>
<td>0.1569</td>
<td>[1e-04, 2.3119]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{12} )</td>
<td>0.0102</td>
<td>[-0.1174, 0.5217]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{22} )</td>
<td>0.0526</td>
<td>[5e-04, 0.7788]</td>
</tr>
</tbody>
</table>

Table 2.17: Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=1.
<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>118.149</td>
<td></td>
<td>$\lambda$</td>
<td>254.9367</td>
<td>[224.1491, 287.884]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$p_1$</td>
<td>0.136</td>
<td>[0.0632, 0.1996]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-115.2929</td>
<td>[-115.6065, -115.2125]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>32.3193</td>
<td>[32.1073, 32.5561]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0256</td>
<td>[0.005, 0.1804]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>0.0051</td>
<td>[-0.0291, 0.0228]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0114</td>
<td>[0.0018, 0.0238]</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>$p_2$</td>
<td>0.1169</td>
<td>[0.0378, 0.2062]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-115.4105</td>
<td>[-115.7096, -114.5274]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>33.3256</td>
<td>[32.0978, 35.9865]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0204</td>
<td>[1e-04, 0.1569]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>0.0032</td>
<td>[-0.0317, 0.0154]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0038</td>
<td>[2e-04, 0.025]</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>$p_3$</td>
<td>0.0965</td>
<td>[0.021, 0.1979]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-115.3292</td>
<td>[-116.2506, -114.613]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>34.5129</td>
<td>[32.6604, 36.0037]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.1063</td>
<td>[1e-04, 0.4414]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>0.1368</td>
<td>[-0.6827, 0.0055]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.319</td>
<td>[2e-04, 1.5015]</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>$p_4$</td>
<td>0.1525</td>
<td>[0.0664, 0.2245]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-115.6458</td>
<td>[-117.3501, -116.0342]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>33.883</td>
<td>[33.2269, 36.0459]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.6178</td>
<td>[0.0438, 1.9118]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>0.6011</td>
<td>[-1.3622, -0.0095]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.9671</td>
<td>[0.0053, 2.409]</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>$p_5$</td>
<td>0.0866</td>
<td>[0.0495, 0.1554]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-117.207</td>
<td>[-117.6817, -116.6336]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>34.8193</td>
<td>[33.7661, 36.1126]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.5234</td>
<td>[0.0279, 2.6779]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>0.1373</td>
<td>[-0.8146, -0.0031]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.1424</td>
<td>[0.003, 1.446]</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>$p_6$</td>
<td>0.2509</td>
<td>[0.1935, 0.3124]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-117.8619</td>
<td>[-118.8611, -117.8553]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>36.3874</td>
<td>[36.3886, 36.3941]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0353</td>
<td>[1e-04, 2e-04]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>0.8022</td>
<td>[0, 0]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0011</td>
<td>[1e-04, 2e-04]</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>$p_7$</td>
<td>0.1606</td>
<td>[0.1221, 0.2043]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-119.8152</td>
<td>[-120.3779, -119.095]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>34.8298</td>
<td>[34.2801, 35.3226]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>2.2508</td>
<td>[1.1521, 4.1948]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-1.8523</td>
<td>[-3.3138, -0.9842]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>2.0317</td>
<td>[1.2398, 3.2675]</td>
</tr>
</tbody>
</table>

Table 2.18: Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=26.
<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>120.9912</td>
<td>1</td>
<td>$\lambda$</td>
<td>227.1133</td>
<td>[198.6886, 258.1012]</td>
</tr>
<tr>
<td></td>
<td>$p_1$</td>
<td></td>
<td>0.2227</td>
<td>[0.1728, 0.2802]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td></td>
<td>-115.1875</td>
<td>[-115.1985, -115.1766]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td></td>
<td>32.4698</td>
<td>[32.4552, 32.4841]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td></td>
<td>0.0018</td>
<td>[0.0012, 0.0027]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td></td>
<td>-0.0021</td>
<td>[-0.0031, -0.0014]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td></td>
<td>0.0031</td>
<td>[0.0021, 0.0044]</td>
</tr>
<tr>
<td>2</td>
<td>$p_2$</td>
<td></td>
<td>0.1797</td>
<td>[0.1309, 0.2581]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td></td>
<td>-115.5312</td>
<td>[-115.8982, -115.4126]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td></td>
<td>32.5752</td>
<td>[32.4121, 32.9356]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td></td>
<td>0.0777</td>
<td>[0.0197, 0.3576]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td></td>
<td>-0.0727</td>
<td>[-0.3472, -0.0124]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td></td>
<td>0.1742</td>
<td>[0.0905, 0.4304]</td>
</tr>
<tr>
<td>3</td>
<td>$p_3$</td>
<td></td>
<td>0.0631</td>
<td>[0.0267, 0.1053]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td></td>
<td>-116.6056</td>
<td>[-116.6466, -116.5644]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td></td>
<td>32.2337</td>
<td>[32.1971, 32.269]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td></td>
<td>0.0015</td>
<td>[3e-04, 0.0051]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td></td>
<td>2e-04</td>
<td>[-0.0013, 0.0021]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td></td>
<td>0.0012</td>
<td>[3e-04, 0.004]</td>
</tr>
<tr>
<td>4</td>
<td>$p_4$</td>
<td></td>
<td>0.1315</td>
<td>[0.072, 0.2349]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td></td>
<td>-117.1574</td>
<td>[-117.1971, -116.9643]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td></td>
<td>34.0018</td>
<td>[33.7751, 34.1866]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td></td>
<td>0.2151</td>
<td>[1e-04, 0.9487]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td></td>
<td>-0.1143</td>
<td>[-0.6212, 0]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td></td>
<td>0.4555</td>
<td>[1e-04, 1.9721]</td>
</tr>
<tr>
<td>5</td>
<td>$p_5$</td>
<td></td>
<td>0.1593</td>
<td>[0.0763, 0.2412]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td></td>
<td>-117.3218</td>
<td>[-117.8453, -117.1582]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td></td>
<td>34.1862</td>
<td>[33.9192, 34.7774]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td></td>
<td>0.4458</td>
<td>[1e-04, 0.947]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td></td>
<td>-0.1765</td>
<td>[-0.6189, 0.1976]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td></td>
<td>1.1384</td>
<td>[1e-04, 2.5333]</td>
</tr>
<tr>
<td>6</td>
<td>$p_6$</td>
<td></td>
<td>0.1467</td>
<td>[0.1114, 0.1868]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td></td>
<td>-117.8532</td>
<td>[-117.8629, -117.8434]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td></td>
<td>36.0493</td>
<td>[36.0424, 36.0564]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td></td>
<td>9e-04</td>
<td>[5e-04, 0.0014]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td></td>
<td>-5e-04</td>
<td>[-8e-04, -3e-04]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td></td>
<td>8e-04</td>
<td>[3e-04, 7e-04]</td>
</tr>
<tr>
<td>7</td>
<td>$p_7$</td>
<td></td>
<td>0.0969</td>
<td>[0.0633, 0.1344]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td></td>
<td>-120.9826</td>
<td>[-121.1669, -120.7924]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td></td>
<td>36.1383</td>
<td>[35.8872, 36.3827]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td></td>
<td>0.1429</td>
<td>[0.0723, 0.281]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td></td>
<td>-0.1298</td>
<td>[-0.2881, -0.0458]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td></td>
<td>0.2457</td>
<td>[0.1217, 0.4874]</td>
</tr>
</tbody>
</table>

Table 2.19: Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=27.
<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>123.9707</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>λ</td>
<td>2356.2806</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>p1</td>
<td>0.0678</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>μ1</td>
<td>-115.0761</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>μ2</td>
<td>32.0818</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>σ11</td>
<td>0.0022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>σ12</td>
<td>-6e-04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>σ22</td>
<td>0.0019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>[0.0412, 0.0967]</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>[-115.1028, -115.0604]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>[ 32.0603, 32.1077 ]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>[ 7e-04, 0.0044 ]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>[-0.0014, 0 ]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>[9e-04, 0.0035]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>p2</td>
<td>0.0847</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>μ1</td>
<td>-115.1883</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>μ2</td>
<td>32.1373</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>σ11</td>
<td>0.0075</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>σ12</td>
<td>-0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>σ22</td>
<td>0.0025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>[0.0467, 0.1533]</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>[-115.2253, -115.1447 ]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>[32.0933, 32.1696 ]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>[ 7e-04, 0.0143 ]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>[-0.0047, 9e-04]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>[0.0013, 0.0044]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>p3</td>
<td>0.1076</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>μ1</td>
<td>-115.2642</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>μ2</td>
<td>32.1952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>σ11</td>
<td>0.0047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>σ12</td>
<td>-0.0029</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>σ22</td>
<td>0.0036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>[0.0532, 0.2534]</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>[-115.3244, -115.1957]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>[32.1373, 32.2542]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>[ 7e-04, 0.0144 ]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>[-0.0068, -1e-04]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>[0.0014, 0.0067]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>p4</td>
<td>0.0843</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>μ1</td>
<td>-115.3084</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>μ2</td>
<td>32.2291</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>σ11</td>
<td>0.0026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>σ12</td>
<td>-0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>σ22</td>
<td>0.0037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>[0.0504, 0.1518]</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>[-115.3724, -115.2754]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>[32.1834, 32.31 ]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>[ 3e-04, 0.011 ]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>[-0.0077, 2e-04 ]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>[0.0012, 0.0078]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>p5</td>
<td>0.0748</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>μ1</td>
<td>-115.3685</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>μ2</td>
<td>32.3106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>σ11</td>
<td>0.0032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>σ12</td>
<td>-0.0021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>σ22</td>
<td>0.0044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>[0.0379, 0.1151]</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>[-115.63, -115.3257]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>[32.2423, 32.5017]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>[ 2e-04, 0.0316 ]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>[+0.02, 4e-04]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>[0.001, 0.019]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>p6</td>
<td>0.0739</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>μ1</td>
<td>-115.4232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>μ2</td>
<td>32.4057</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>σ11</td>
<td>0.0038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>σ12</td>
<td>-0.0025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>σ22</td>
<td>0.0088</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>[0.0051, 0.1254]</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>[-115.6445, -115.3298]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>[32.2565, 32.9714]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>[ 2e-04, 0.0188 ]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>[-0.0152, 0.0048]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>[0.0011, 0.0693]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>p7</td>
<td>0.0716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>μ1</td>
<td>-115.5736</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>μ2</td>
<td>32.4805</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>σ11</td>
<td>0.0098</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>σ12</td>
<td>-0.0057</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>σ22</td>
<td>0.0101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>[0.0065, 0.2333]</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>[-115.7684, -115.3654]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>[32.2936, 32.963]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>[ 5e-04, 0.0499 ]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>[-0.0313, 0.0062]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>[0.001 , 0.0683]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>p8</td>
<td>0.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>μ1</td>
<td>-115.6935</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>μ2</td>
<td>32.5443</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>[0.0267, 0.1942]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>[-115.8513, -115.5659]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>[32.4116, 32.6944]</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>p9</td>
<td>0.0702</td>
<td>[0.0314, 0.1729]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-115.7606</td>
<td>[-115.8668, -115.6924]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>32.6184</td>
<td>[32.5406, 32.6967]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>0.0064</td>
<td>[3e-04, 0.106]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.0045</td>
<td>[-0.0843, 0]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.0041</td>
<td>[2e-04, 0.0668]</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>p10</td>
<td>0.0666</td>
<td>[0.0056, 0.1012]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-115.7921</td>
<td>[-115.9593, -115.7491]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>32.6475</td>
<td>[32.6084, 33.0994]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>0.0097</td>
<td>[2e-04, 0.1398]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.0075</td>
<td>[-0.1144, 0]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.0066</td>
<td>[1e-04, 0.0943]</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>p11</td>
<td>0.0643</td>
<td>[0.0039, 0.1094]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-115.8113</td>
<td>[-116.1001, -115.7589]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>32.708</td>
<td>[32.61, 33.1484]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>0.0087</td>
<td>[2e-04, 0.1274]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.0064</td>
<td>[-0.1021, 0]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.0057</td>
<td>[2e-04, 0.084]</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>p12</td>
<td>0.0579</td>
<td>[0.0026, 0.1036]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-115.8782</td>
<td>[-116.1551, -115.7851]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>32.8119</td>
<td>[32.6648, 33.7946]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>0.0114</td>
<td>[4e-04, 0.1459]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.0081</td>
<td>[-0.1152, 0]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.0071</td>
<td>[2e-04, 0.0964]</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>p13</td>
<td>0.0352</td>
<td>[0.0031, 0.0645]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-115.9485</td>
<td>[-116.4783, -115.8375]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>32.834</td>
<td>[32.6711, 33.7536]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>0.0126</td>
<td>[7e-04, 0.1209]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.0069</td>
<td>[-0.0972, -2e-04]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.0071</td>
<td>[3e-04, 0.0931]</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>p14</td>
<td>0.0234</td>
<td>[0.0024, 0.0496]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-116.2336</td>
<td>[-117.967, -115.8552]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>33.0968</td>
<td>[32.6881, 33.802]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>0.1175</td>
<td>[8e-04, 0.234]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.0651</td>
<td>[-0.1669, -1e-04]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.0813</td>
<td>[4e-04, 0.2107]</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>p15</td>
<td>0.0169</td>
<td>[0.0039, 0.0399]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-117.9259</td>
<td>[-118.3668, -117.6523]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>35.4991</td>
<td>[33.8817, 36.2636]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>1.0666</td>
<td>[1e-04, 4.4516]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.6123</td>
<td>[-2.6371, 0]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.6016</td>
<td>[1e-04, 2.5043]</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>p16</td>
<td>0.0113</td>
<td>[0.0033, 0.0267]</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2.20: Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=28.

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td></td>
<td>$\lambda$</td>
<td>472.9284</td>
<td>430.3604, 516.7343</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$p_1$</td>
<td>0.0335</td>
<td>[0.0064, 0.0833]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-120.352</td>
<td>[-120.5209, -120.1918]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>33.8086</td>
<td>[33.7348, 33.8813]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.03</td>
<td>[0.0022, 0.0801]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0115</td>
<td>[-0.0317, -8e-04]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>1.2692</td>
<td>[8e-04, 1.055]</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>$p_2$</td>
<td>0.1832</td>
<td>[0.1282, 0.3189]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-115.8042</td>
<td>[-115.8222, -115.7812]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>32.6554</td>
<td>[32.6366, 32.6693]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0086</td>
<td>[0.0025, 0.0056]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0028</td>
<td>[-0.0037, -0.0017]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0048</td>
<td>[0.0013, 0.0029]</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>$p_3$</td>
<td>0.5196</td>
<td>[0.3919, 0.6026]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-115.2858</td>
<td>[-115.3018, -115.2691]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>32.2204</td>
<td>[32.2072, 32.2332]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0231</td>
<td>[0.0152, 0.0217]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0135</td>
<td>[-0.016, -0.0109]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.0138</td>
<td>[0.0095, 0.0138]</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>$p_4$</td>
<td>0.093</td>
<td>[0.0618, 0.1287]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-117.6363</td>
<td>[-118.1537, -117.1166]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>34.3654</td>
<td>[33.9533, 34.7807]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>2.4702</td>
<td>[1.5388, 3.8941]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-1.1341</td>
<td>[-2.0155, -0.5249]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{22}$</td>
<td>1.4521</td>
<td>[0.9067, 2.2768]</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>$p_5$</td>
<td>0.0317</td>
<td>[0.0054, 0.0804]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_1$</td>
<td>-117.608</td>
<td>[-117.7211, -117.5082]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_2$</td>
<td>32.2132</td>
<td>[32.1375, 32.2895]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.0131</td>
<td>[7e-04, 0.0272]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.0048</td>
<td>[-0.0164, 2e-04]</td>
</tr>
</tbody>
</table>

Continued on next page
### Table 2.21: Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=29.

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>181.5083</td>
<td>λ</td>
<td>300.9285</td>
<td>[267.0044, 335.5887]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>p1</td>
<td>0.1669</td>
<td>[0.1233, 0.2169]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-115.3954</td>
<td>[-115.4576, -115.3396]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>32.6313</td>
<td>[32.5088, 32.767]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>0.04</td>
<td>[0.0255, 0.0651]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.0841</td>
<td>[-0.1371, -0.0532]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.1903</td>
<td>[0.1224, 0.3054]</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>p2</td>
<td>0.1284</td>
<td>[0.0628, 0.2033]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-116.393</td>
<td>[-116.5929, -116.2221]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>33.702</td>
<td>[33.3406, 34.0735]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>0.1797</td>
<td>[0.0758, 0.3534]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.1503</td>
<td>[-0.3136, -0.0362]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.4018</td>
<td>[0.0249, 0.9269]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>p3</td>
<td>0.0856</td>
<td>[0.0271, 0.1701]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-116.6294</td>
<td>[-117.1297, -116.3627]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>33.7162</td>
<td>[33.1997, 34.1393]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>0.1812</td>
<td>[0.0189, 0.3677]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.1254</td>
<td>[-0.2912, -0.0057]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.1714</td>
<td>[0.0122, 0.8383]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>p4</td>
<td>0.0488</td>
<td>[0.0087, 0.1086]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-117.6643</td>
<td>[-118.2598, -117.3626]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>33.0357</td>
<td>[32.0635, 34.0689]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>0.4328</td>
<td>[0.0114, 2.1869]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.2712</td>
<td>[-1.3569, -0.0021]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.2347</td>
<td>[0.006, 1.014]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>p5</td>
<td>0.0975</td>
<td>[0.0512, 0.2158]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-117.8785</td>
<td>[-118.3491, -117.6023]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>35.1848</td>
<td>[33.4995, 36.0489]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ11</td>
<td>0.3286</td>
<td>[0.0247, 2.6262]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ12</td>
<td>-0.117</td>
<td>[-0.5871, 0.0075]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>σ22</td>
<td>0.3295</td>
<td>[0.0077, 1.6265]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>p6</td>
<td>0.0982</td>
<td>[0.0448, 0.2461]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ1</td>
<td>-118.1485</td>
<td>[-118.5841, -117.6336]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ2</td>
<td>34.7802</td>
<td>[33.7694, 36.0456]</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.22: Model 2 (Earthquakes application): Posterior estimates of the model parameters, regarding the PME Poisson intensity surface at time index=44.

<table>
<thead>
<tr>
<th>Time</th>
<th>Component</th>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.3515</td>
<td>[0.0253, 2.5232]</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.1491</td>
<td>[-0.6024, 0.0111]</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.3897</td>
<td>[0.0084, 1.5274]</td>
</tr>
<tr>
<td>7</td>
<td>$p_7$</td>
<td>0.17</td>
<td></td>
<td>[0.0575, 0.2814]</td>
</tr>
<tr>
<td>7</td>
<td>$\mu_1$</td>
<td>-118.4607</td>
<td></td>
<td>[-119.0184, -117.6987]</td>
</tr>
<tr>
<td>7</td>
<td>$\mu_2$</td>
<td>34.9134</td>
<td></td>
<td>[33.9645, 36.0173]</td>
</tr>
<tr>
<td>7</td>
<td>$\sigma_{11}$</td>
<td>1.6429</td>
<td></td>
<td>[0.0375, 3.0936]</td>
</tr>
<tr>
<td>7</td>
<td>$\sigma_{12}$</td>
<td>-0.0493</td>
<td></td>
<td>[-0.5944, 0.6091]</td>
</tr>
<tr>
<td>7</td>
<td>$\sigma_{22}$</td>
<td>0.9385</td>
<td></td>
<td>[0.0128, 1.7249]</td>
</tr>
<tr>
<td>8</td>
<td>$p_8$</td>
<td>0.0627</td>
<td></td>
<td>[0.0347, 0.0942]</td>
</tr>
<tr>
<td>8</td>
<td>$\mu_1$</td>
<td>-120.3736</td>
<td></td>
<td>[-120.5867, -120.1772]</td>
</tr>
<tr>
<td>8</td>
<td>$\mu_2$</td>
<td>36.3105</td>
<td></td>
<td>[36.1423, 36.4771]</td>
</tr>
<tr>
<td>8</td>
<td>$\sigma_{11}$</td>
<td>0.1333</td>
<td></td>
<td>[0.0591, 0.2982]</td>
</tr>
<tr>
<td>8</td>
<td>$\sigma_{12}$</td>
<td>-0.1066</td>
<td></td>
<td>[-0.2272, -0.0474]</td>
</tr>
<tr>
<td>8</td>
<td>$\sigma_{22}$</td>
<td>0.0923</td>
<td></td>
<td>[0.0432, 0.1929]</td>
</tr>
<tr>
<td>9</td>
<td>$p_9$</td>
<td>0.0401</td>
<td></td>
<td>[0.0146, 0.0731]</td>
</tr>
<tr>
<td>9</td>
<td>$\mu_1$</td>
<td>-121.1582</td>
<td></td>
<td>[-121.3208, -120.9182]</td>
</tr>
<tr>
<td>9</td>
<td>$\mu_2$</td>
<td>35.6884</td>
<td></td>
<td>[35.5934, 35.7983]</td>
</tr>
<tr>
<td>9</td>
<td>$\sigma_{11}$</td>
<td>0.0471</td>
<td></td>
<td>[0.0139, 0.1617]</td>
</tr>
<tr>
<td>9</td>
<td>$\sigma_{12}$</td>
<td>-0.013</td>
<td></td>
<td>[-0.0485, 0.026]</td>
</tr>
<tr>
<td>9</td>
<td>$\sigma_{22}$</td>
<td>0.0163</td>
<td></td>
<td>[0.005, 0.047]</td>
</tr>
<tr>
<td>10</td>
<td>$p_{10}$</td>
<td>0.1018</td>
<td></td>
<td>[0.0728, 0.1352]</td>
</tr>
<tr>
<td>10</td>
<td>$\mu_1$</td>
<td>-121.3327</td>
<td></td>
<td>[-121.4018, -121.2597]</td>
</tr>
<tr>
<td>10</td>
<td>$\mu_2$</td>
<td>36.6894</td>
<td></td>
<td>[36.6407, 36.7383]</td>
</tr>
<tr>
<td>10</td>
<td>$\sigma_{11}$</td>
<td>0.0354</td>
<td></td>
<td>[0.021, 0.0609]</td>
</tr>
<tr>
<td>10</td>
<td>$\sigma_{12}$</td>
<td>-0.0161</td>
<td></td>
<td>[-0.0302, -0.0075]</td>
</tr>
<tr>
<td>10</td>
<td>$\sigma_{22}$</td>
<td>0.0168</td>
<td></td>
<td>[0.0097, 0.0283]</td>
</tr>
</tbody>
</table>
2.C Figures

Figure 2.1: Model 1 (Simulation): Displaying the true Hawkes process intensity function (red), along with the estimated Hawkes process based on the posterior means (black) of the model parameters. We also present 95% credible sets for the Hawkes process based on the bounds of the 95% credible sets of the parameters. The temporal point pattern is shown at the bottom of the figure using red line segments.
Figure 2.2: Model 1 (Simulation): Trace plots, histograms and ACF’s for \( \mu \), \( \alpha \) and \( \beta \). The green dashed line represents the true value and the red line represents the posterior mean for the respective parameters.
Figure 2.3: Model 1 (Simulation): 2-d plots of the true Poisson intensity surfaces (left column) and the corresponding PME intensity surfaces (right column) at times $t_1 = 2.6522$, $t_{13} = 25.3369$ and $t_{27} = 47.6197$. The point patterns are also shown and the gray crosses represent the mixture component means.
Figure 2.4: Model 1 (Simulation): 3-d plots of the true Poisson intensity surfaces (left column) and the corresponding PME intensity surfaces (right column) at times $t_1 = 2.6522$, $t_{13} = 25.3369$ and $t_{27} = 47.6197$. 
Figure 2.5: Model 2 (Simulation): Displaying the true Hawkes process intensity function (red), along with the estimated Hawkes process based on the posterior means (black) of the model parameters. We also present 95% credible sets for the Hawkes process based on the bounds of the 95% credible sets of the parameters. The temporal point pattern is shown at the bottom of the figure using red line segments.
Figure 2.6: Model 2 (Simulation): Trace plots, histograms and ACF’s for $\mu$, $\alpha$ and $\beta$. The green dashed line represents the true value and the red line represents the posterior mean for the respective parameters.
Figure 2.7: Model 2 (Simulation): 2-d plots of the true Poisson intensity surfaces (left column) and the corresponding PME intensity surfaces (right column) at times $t_1$, $t_{13}$ and $t_{27}$. The point patterns are also shown and the gray crosses represent the mixture component means.
Figure 2.8: Model 2 (Simulation): 3-d plots of the true Poisson intensity surfaces (left column) and the corresponding PME intensity surfaces (right column) at times $t_1$, $t_{13}$ and $t_{27}$. 
Figure 2.9: Earthquakes application: Displaying the time events of the Hawkes process. Note the spike in event arrivals as a result of the major earthquake at Baja, California, on April 4, 2010.

Figure 2.10: Model 1 (Earthquakes application): Displaying the estimated Hawkes process based on the posterior means of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments.
Figure 2.11: Model 1 (Earthquakes application): Trace plots, histograms and ACF’s for $\mu$, $\alpha$ and $\beta$. The red dashed line indicates the posterior means for the model parameters, $\mu$, $\alpha$ and $\beta$. 
Figure 2.12: Model 1 (Earthquakes application): 2-d plots posterior means Poisson intensity surfaces at times $t_1$, $t_{26}$, $t_{27}$, $t_{28}$, $t_{29}$, and $t_{44}$. The point patterns are also shown and the yellow crosses represent the mixture component means.
Figure 2.13: Model 1 (Earthquakes application): 3-d plots of the true Poison intensity surfaces (left column) and the corresponding PME Poison intensity surfaces (right column) at times $t_1$, $t_{26}$, $t_{27}$, $t_{28}$, $t_{29}$, and $t_{44}$. 
Figure 2.14: Model 2 (Earthquakes application): Displaying the estimated Hawkes process based on the posterior means of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments.
Figure 2.15: Model 2 (Earthquakes application): Trace plots, histograms and ACF’s for $\mu$, $\alpha$ and $\beta$. The red dashed line indicates the posterior means for the model parameters, $\mu$, $\alpha$ and $\beta$. 
Figure 2.16: Model 2 (Earthquakes application): 2-d plots PME Poisson intensity surfaces at times $t_1$, $t_{26}$, $t_{27}$, $t_{28}$, $t_{29}$, and $t_{44}$. The point patterns are also shown and the yellow crosses represent the component means.
Figure 2.17: Model 2 (Earthquakes application): 3-d plots of the true Poisson intensity surfaces (left column) and the corresponding PME Poisson intensity surfaces (right column) at times $t_1$, $t_{26}$, $t_{27}$, $t_{28}$, $t_{29}$, and $t_{44}$. 
Chapter 3

Spatial-Temporal Poisson Point Process Models via Conditioning on Location

3.1 Introduction

Consider a point pattern that arises in animal ecology where we place cameras at some locations over an area of interest. Note that the locations may not be selected randomly, however, we will assume that this is the case for the sake of presentation. The cameras take snap-shots only at (random) times when they detect movement across their field of view, i.e., an animal passes by and the camera takes a picture.

In order to model this type of point pattern, we need to consider a modification of the conditional approach of equation (1.13), where the joint intensity function of the marked space-time NHPPP is broken down into three parts via conditioning on the location component of the event. Therefore, we first condition on the location of the camera and require a spatial point process model and then on the time an event occurred and require a temporal model at the given location.

Since many of the parameters of the marked temporal process are conditioned on the location, their a priori models need to be defined over all possible locations in the window of observation, i.e., their priors are modeled via spatial random fields. For an excellent theoretical treatment of random fields see Spodarev (2013).

This chapter is organized as follows. First we introduce spatial-temporal point process
via conditioning on the location component of the event, and then discuss specific choices for the conditional intensity functions involved. We illustrate our modeling approach via a simulated example.

3.2 Modeling a spatial-temporal point process via conditioning on location

First, note that the point patterns under consideration are of the form

\[
\varphi_n = \{(t_k(s_l), s_l), \ k = 1, \ldots, n(s_l), \ l = 1, \ldots, L, \\
s_l \in \mathcal{W}, \ t_k(s_l) \in \mathcal{T} = [0, T),
\]

(3.1)

where \(n(s_l)\) denotes the number of temporal events at location \(s_l, \ l = 1, \ldots, L\). Note that \(n = \sum_{l=1}^{L} n(s_l)\), is the total number of events.

In order to build the space-time model via conditioning on location, we begin by considering a NHPPP model on the locations \(\psi_L = \{s_1, \ldots, s_L\}\), and then given \(\psi_L\), we model temporal processes \(\varphi^{s_l} = \{s_l, t_k(s_l)\}, \ k = 1, \ldots, n(s_l)\}, \ l = 1, \ldots, L\, as conditionally independent marked Hawkes processes. That is, the joint distribution of the point pattern of equation (3.1) is modeled using

\[
f(\varphi_n|\theta) = \left[ \prod_{l=1}^{L} f(\varphi^{s_l}|s_l, \theta(s_l)) \right] f(\psi_L, L),
\]

(3.2)

and we discuss the specific choices for \(f(\psi_L, L)\) and \(f(\varphi^{s_l}|s_l, \theta(s_l))\) next.
3.2.1 Modeling the event locations

We consider the locations $\psi = \{s_1, \ldots, s_L\}$ as a realization of a NHPPP over $W$, with intensity function $\lambda_0(s|\lambda_0, \theta_0)$ given by

$$\lambda_0(s|\lambda_0, \theta_0) = \lambda_0 \varphi_j(s|\mu_j, \Sigma_j),$$

where $\theta_0 = \{m, p, \{\mu_j\}_{j=1}^m, \{\Sigma_j\}_{j=1}^m\}$, $m$ the number of components, $\mu_j$ the $j^{th}$ component mean vector, $\Sigma_j$ the $j^{th}$ component covariance matrix, $p = (p_1, \ldots, p_m)$, $p_j$ the $j^{th}$ component probability, and $\varphi_j(s|\mu_j, \Sigma_j)$ denotes the bivariate normal density $N_2(\mu_j, \Sigma_j)$, $j = 1, 2, \ldots, m$. Note that $\lambda(s|\theta_0)$ is a proper density over $W$ provided that all of the mixture component distributions have their mass in $W$.

In order to remove the intractability of the mixture model, we appeal to data augmentation so that (3.3) becomes the data-augmented intensity

$$\lambda_0(s_l, z_l|\lambda_0, \theta_0) = \lambda_0 \prod_{j=1}^m p_j^{z_{lj}} [\varphi_j(s_l|\mu_j, \Sigma_j)^{z_{lj}}],$$

where $z_l = [z_{l1}, \ldots, z_{lm}]$ are the membership allocation variables for the location $s_l$, $l = 1, 2, \ldots, L$. As a result, the (data-augmented) joint distribution $f(\psi_L, L|\theta_1)$ of the locations and their number $L$, is given by

$$f(\psi_L, z, L|\lambda_0, \theta_1) = \frac{e^{-\lambda_0 L}}{L!} \prod_{l=1}^L \prod_{j=1}^m p_j^{z_{lj}} [\varphi_j(s_l|\mu_j, \Sigma_j)^{z_{lj}}],$$

where $z = \{z_l, l = 1, 2, \ldots, L\}$ denotes all the data-augmentation vectors.

3.2.2 Hawkes point processes

Defining marked Hawkes models via conditioning is straightforward, following similar steps as in the case of the marked NHPPP. More precisely, the point pattern in this case is
of the form
\[ \varphi_n = \{(t_i), \ i = 1, \ldots, n, \ t_i \in \mathcal{W}_i, \} \] (3.6)

and it can be modeled as a marked Hawkes process using a joint intensity function \( \lambda(t, \xi) \).
Depending on the type of data under consideration, we condition either on the mark value or the time stamp of the event. That is, we can model the joint intensity function using

\[ \lambda(t|\theta) = \lambda \times \lambda_1(t|\theta), \] (3.7)

where \( \lambda_1 \) is a proper densities and \( \theta \) denotes all the parameters of the STPP model. Thus the joint distribution

\[
f(\varphi_n, n|\theta) = \frac{e^{-\Lambda(\mathcal{W}_t)}}{n!} \prod_{i=1}^{n} \lambda(t_i|\theta)
= \frac{\lambda^n e^{-\lambda}}{n!} \prod_{i=1}^{n} \lambda(t_i|\theta)),
\] (3.8)

since

\[ \Lambda(\mathcal{W}_t) = \int_{\mathcal{W}_t} \lambda(t|\theta) dt = \lambda, \]

where \( \theta \) denotes all the parameters of the Hawkes process model.

In particular, we consider a (normalized) Hawkes process for the \( \lambda(t|\theta) \) given by

\[ \lambda_1(t|\theta_1) = \frac{1}{\lambda(\mu, \alpha, \beta, T)} h(t|\mu, \alpha, \beta, \mathcal{H}_t), \] (3.10)

where \( \theta = (\mu, \alpha, \beta) \), and \( h(t|\mu, \alpha, \beta, \mathcal{H}_t) \) is given in equation \( \text{[2.6]} \).

As a result, the joint distribution of the Hawkes process is given by

\[
f(\varphi_n, n|\theta) = \frac{\lambda^n e^{-\lambda}}{n!} \prod_{i=1}^{n} \lambda_1(t_i|\theta)
= \frac{\lambda^n e^{-\lambda}}{n!} \times \prod_{i=1}^{n} \frac{1}{\lambda(\mu, \alpha, \beta, T)} h(t_i|\mu, \alpha, \beta, \mathcal{H}_t),
\] (3.11)

\[
= \frac{\lambda^n e^{-\lambda}}{n!} \times \prod_{i=1}^{n} \frac{1}{\lambda(\mu, \alpha, \beta, T)} h(t_i|\mu, \alpha, \beta, \mathcal{H}_t),
\] (3.12)
and therefore,
\[ f(\varphi_n, n|\theta) = \frac{\lambda^n e^{-\lambda}}{n!} \left( \frac{1}{\Lambda(\mu, \alpha, \beta, t)} \right)^n \times \prod_{i=1}^{n} \left( \mu + \alpha \beta \sum_{j: t_j < t_i} e^{-\beta(t_i-t_j)} \right), \quad (3.13) \]
where \( \theta = (\lambda, \theta_1) \) denotes all the parameters of the model, \( \theta_1 = (\mu, \alpha, \beta) \).

In view of (3.2), the Hawkes process of equation (3.13) corresponds to the term \( f(\varphi^s_i|s_l, \theta(s_l)) \), and will be used to model the marked temporal process at each given location \( s_l \). In particular, we have
\[ f(\varphi^s_i|s_l, \theta(s_l)) = \frac{\lambda(s_l)^n(s_i) e^{-\lambda(s_l)}}{n(s_l)!} \prod_{i=1}^{n(s_l)} \lambda_1(t_i(s_l)|s_l, \theta_1(s_l)), \quad (3.14) \]
and we notice that all the parameters of model (3.13) are now spatially varying, given the location \( s_l \), \( l = 1, 2, \ldots, L \).

A natural approach to modeling the term \( \lambda_1(t(s)|s, \theta_1(s)) \) appearing in the joint (3.14), is via a Hawkes process on \( T = [0, T] \) with parameters depending on location; that is, we set
\[ \lambda_1(t(s)|s, \theta_1(s)) = h(t(s)|s, \theta_1(s), \mathcal{H}_{t(s)}) = \mu(s) + \alpha(s) \beta(s) \sum_{i: t_i < t(s)} e^{-\beta(s)(t(s)-t_i)}, \quad (3.15) \]
where \( \theta_1(s) = (\mu(s), \alpha(s), \beta(s)) \), \( \mathcal{H}_{t(s)} \) is the history of the process, and at each location \( s \), we have \( \mu(s) > 0 \) is the initial intensity at time \( t(s) = 0 \), and \( \alpha(s) \in (0, 1) \) is the size of the self exciting jumps relative to the constant exponential decay rate \( \beta(s) > 0 \). The point pattern we model using (3.15) is of the form \( \psi_n = \{(s_1, t_1(s_1)), \ldots, (s_n, t_n(s_n))\} \). Note that if we choose model (3.15), then the term \( \lambda \) appearing in (3.7) is absorbed in the parameters of the Hawkes process \( \theta_1(s) \) and we do not need to normalize \( \lambda_1(t(s)|s, \theta_1(s)) \). However, if we want to keep the term \( \lambda \) in (3.7), then in order to make \( \lambda_1(t(s)|s, \theta_1(s)) \) a proper density, we set
\[ \lambda_1(t(s)|s, \theta_1(s)) = \frac{1}{\Lambda(s, \mu(s), \alpha(s), \beta(s), T)} h(t(s)|s, \mu(s), \alpha(s), \beta(s), \mathcal{H}_{t(s)}), \quad (3.16) \]
where

\[ \Lambda(s, T) = \Lambda(s, \mu(s), \alpha(s), \beta(s), T) = \int_0^T h(t(s)|s, \mu(s), \alpha(s), \beta(s), \mathcal{H}_t(s)) \, dt \]
\[ = \mu(s) T + \alpha(s) \sum_{i:t_i < t} (1 - e^{-\beta(s)(T-t_i)}) \, dt. \]

3.3 Hierarchical Bayesian Modeling

We present the general hierarchical Bayesian formulation (see Section 2.3) next.

3.3.1 Stage 1: Data Model

Based on (3.2), (3.5) and (3.14), the data model is given by

\[
f(\phi_n, n, z|\theta) = \frac{e^{-\lambda_0 L}}{L!} \prod_{l=1}^{L} \prod_{j=1}^{m} p_j^{z_{i,l}} |2\pi \Sigma_j|^{-z_{i,l}/2} e^{-z_i/2 (s_l-\mu_j)^T \Sigma_j^{-1} (s_l-\mu_j)}
\]

\[ \times \prod_{l=1}^{L} e^{-\Lambda(s_l, T)} \prod_{k=1}^{n(s_l)} \left( \mu(s_l) + \alpha(s_l) \beta(s_l) \sum_{i:t_i < t_k(s_l)} e^{-\beta(s_l)(t_k(s_l)-t_i)} \right) \]

where \( \theta \) denotes all the model parameters, with \( \Lambda(s_l, T) = \int_{t=0}^{T} h(t|s_l, \mu(s_l), \alpha(s_l), \beta(s_l), \mathcal{H}_t) \, dt \).

Our modeling approach requires non-dynamic modeling of the average number of locations \( \lambda_0 \), and the mixture parameters \( \theta_0 = \{ m, p, \{ \mu_j \}_{j=1}^{m}, \{ \Sigma_j \}_{j=1}^{m} \} \), for the spatial process. Then, conditional on the spatial location \( s \), the temporal process is modeled using a Hawkes process with parameters \( \mu(s), \alpha(s), \) and \( \beta(s) \) via equation (3.15).

Notice that the temporal processes are spatially varying modeled with different Hawkes processes. Thus we model the hawkes parameters \( \mu(s), \alpha(s), \) and \( \beta(s) \) that drive the Hawkes process at each location. and since \( \mu(s) > 0, \alpha(s) \in (0,1), \) and \( \beta(s) > 0, \) for any \( s \in \mathcal{W} \) are spatially varying we consider the following models for the Hawkes parameters

\[
\mu^*(s) = \log \mu(s) \sim N(\zeta_\mu, \sigma_\mu^2),
\]

(3.18)
\[ \alpha^*(s) = \log \left( \frac{\alpha(s)}{1 - \alpha(s)} \right) \sim N(\zeta_\alpha, \sigma_\alpha^2), \]  

and

\[ \beta^*(s) = \log \beta(s) \sim N(\zeta_\beta, \sigma_\beta^2), \]  

It can be shown that

\[ f(\mu(s|\zeta_\mu, \sigma^2_\mu)) = \frac{1}{\mu(s)\sigma_\mu\sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma^2_\mu} \left( \log(\mu(s)) - \zeta_\mu \right)^2 \right) \]  

\[ f(\alpha(s|\zeta_\alpha, \sigma^2_\alpha)) = \frac{1}{\alpha(s)(1 - \alpha(s))\sigma_\alpha\sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma^2_\alpha} \left( \log \left( \frac{\alpha(s)}{1 - \alpha(s)} \right) - \zeta_\alpha \right)^2 \right) \]  

\[ f(\beta(s|\zeta_\beta, \sigma^2_\beta)) = \frac{1}{\beta(s)\sigma_\beta\sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma^2_\beta} \left( \log(\beta(s)) - \zeta_\beta \right)^2 \right) \]  

where \( \zeta_\mu, \zeta_\alpha, \zeta_\beta \in \mathcal{R} \), and \( \sigma^2_\mu > 0, \sigma^2_\alpha > 0, \) and \( \sigma^2_\beta > 0 \) denote the parameters.

### 3.3.2 Stage 2: Priors for the model parameters

The prior for \( \lambda_0 \) is simply \( \lambda_0|a_\lambda, b_\lambda \sim G(a_\lambda, b_\lambda) \). We discuss below the priors for the parameters of the location process and the Hawkes process.

**Prior for the location process parameters**

For the mixture parameters involved in the NHPPP of locations, we model *a priori* \( p|d \sim Dirichlet(d) \), for some fixed \( d = (d_1, \ldots, d_m) \), and

\[ \mu_j \sim N_2(\nu, K^{-1}), \]  

\[ \Sigma_j|\mathbf{B} \sim IWishart_2(2a, 2\mathbf{B}), \]  

or equivalently

\[ \Sigma_j^{-1}|\mathbf{B} \sim Wishart_2(2a, (2\mathbf{B})^{-1}), \]
for all \( j = 1, 2, \ldots, m \), and we add an additional stage in the hierarchy in order to model the hyper-parameter matrix \( B \) as follows

\[
B \sim \text{Wishart}_2(2g, (2H)^{-1}),
\]

with \( a > 0 \), \( g > 0 \), \( \nu \), \( K \), and \( H \), known. These hyper-parameters are selected appropriately in order to accomplish several tasks, including guaranteeing propriety of the posterior distribution, providing good proposals in order to sample from likely values of the parameters and explore the parameter space (empirical Bayes priors), speed up each iteration of the MCMC, as well as, tuning the MCMC algorithm. In particular, we choose \( a = 1 \), \( b = 1 \), \( a = 3 \), \( g = .3 \), \( d = 1 \), \( \nu = \frac{1}{L} \sum_{l=1}^{L} s_l \), \( K = 100\text{diag}(R_1^2, R_2^2) \), \( H = 100 \text{ diag}(g/(aR_1^2), g/(aR_2^2)) \), where \( R_1 \) and \( R_2 \) the range of values for the \( x \) and \( y \) coordinate, respectively, of the observed spatial locations \( \{s_1, \ldots, s_L\} \). The setup for the BHM above follows from the papers of Richardson and Green (1997), Stephens (2000) and Micheas (2014).

**Spatial processes for the Hawkes process parameters**

For the the hawkes parameters involved in the temporal processes, we model \( \zeta_{\mu} \sim N(\mu_{\zeta_{\mu}}, \sigma_{\zeta_{\mu}}^2) \), \( \zeta_{\alpha} \sim N(\mu_{\zeta_{\alpha}}, \sigma_{\zeta_{\alpha}}^2) \), \( \zeta_{\beta} \sim N(\mu_{\zeta_{\beta}}, \sigma_{\zeta_{\beta}}^2) \), \( \sigma_{\mu^2} \sim \text{IG}(a_{\sigma_{\mu^2}}, b_{\sigma_{\mu^2}}) \), \( \sigma_{\alpha^2} \sim \text{IG}(a_{\sigma_{\alpha^2}}, b_{\sigma_{\alpha^2}}) \), and \( \sigma_{\beta^2} \sim \text{IG}(a_{\sigma_{\beta^2}}, b_{\sigma_{\beta^2}}) \).

### 3.3.3 Posterior distribution

Let \( \theta = (\lambda_0, \theta_0, \theta_h) \), \( \theta_0 = \{m, p, \{\mu_j\}_{j=1}^m, \{\Sigma_j\}_{j=1}^m\} \), and \( \theta_h = \{\{\mu(s_l), \alpha(s_l), \beta(s_l)\}_{l=1}^L\} \), \( \zeta_{\mu}, \zeta_{\alpha}, \zeta_{\beta}, \sigma_{\mu^2}, \sigma_{\alpha^2}, \sigma_{\beta^2} \), \( \sigma_{\mu^2} \), \( \sigma_{\alpha^2} \), \( \sigma_{\beta^2} \), denote the model and process parameters and hyper-parameters that are treated as random and require prior distributions. The full posterior distribution is given by

\[
\pi(\theta, z|\varphi_{\xi,n}, n) \propto f(\varphi_{\xi,n}, n, z|\lambda_0, \theta_0)\pi(\lambda_0, \theta_0, \theta_h)\pi(z),
\] (3.26)
where \( f(\varphi_{\xi,n}, n, z|\theta) \) is given by (3.17), and the priors are given in Section 3.3.2. The full-
conditional distributions are presented in Appendix 3.A.

3.3.4 Simulation of the space-time model conditioning on location

Next we present the general algorithm that is used to simulate space-time data from the
proposed model of equation (3.17).

1. Set the finite mixture model parameters that control the spatial randomness, that is,
the average number of locations \( \lambda_0 \) and \( \theta_0 = \{ m, \mathbf{p}, \{ \mu_j \}_{j=1}^m, \{ \Sigma_j \}_{j=1}^m \} \).

2. Set the parameters that control the temporal randomness (hawkes processes), that is
\( \zeta_\mu, \zeta_\alpha, \zeta_\beta, \sigma^2_\mu, \sigma^2_\alpha, \) and \( \sigma^2_\beta \)

3. Using the finite mixture parameters set in step 1, simulate a point pattern from a
NHPPP with intensity function given in equation (3.3).

4. For each spatial location simulate the Hawkes process parameters \( \mu(s_l), \alpha(s_l), \) and
\( \beta(s_l), \) for \( l = 1, \ldots, L, \) using the parameters obtained in step 2.

5. For a given time interval \( T = [0, T) \) and the Hawkes process parameters obtained
in step 4, simulate realizations \( \psi_{K(s_l)}(s_l) = \{ t_1(s_l), \ldots, t_{K(s_l)}(s_l) \} \sim Hawkes(\mu(s_l), \alpha(s_l), \beta(s_l), T), \) for \( l = 1, \ldots, L, \) using Algorithm 2 of Ogata (1988). The simulated
temporal values \( \psi_{K(s_l)} \) represent the occurrence times of events at location \( s_l \) over the
temporal window \( [0, T) \) and constitute the “history” of the Hawkes process at location
\( s_l, \) denoted by \( \mathcal{H}_{t(s_l)}. \)

6. The space-time point pattern is given by \( \varphi_n = \{ (s_l, t_k(s_l)) \}, \ k = 1, \ldots, K(s_l), \ l = \)
1, \ldots, L\}. 

125
3.4 Simulated example

First we generate the space-time data based on the algorithm presented in Section 3.3.4 with the model parameters being known (true values) and then obtain their estimates using the proposed BHM methodology. In order to assess the performance of the models we then compare the estimates with the true values of the parameters used to simulate the data.

3.4.1 True parameter values and the simulation procedure

The simulated data was obtained using the procedure outlined in Section 3.3.4 as follows:

1. Set the finite mixture parameters that control the spatial randomness to $\lambda_0 = 200$, $m = 3$, $p = [.5, .3, .2]$, $\mu_1 = [7, 4]^T$, $\mu_2 = [5, 8]^T$, $\mu_3 = [3, 3]^T$, $\Sigma_1 = \begin{pmatrix} 0.5 & -0.1 \\ -0.1 & 0.5 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and $\Sigma_3 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$.

2. Set the parameters $\mu(s), \alpha(s)$ and $\beta(s)$, that is $\zeta_\mu = 0$, $\zeta_\alpha = 0$, $\zeta_\beta = 0$, $\sigma_\mu^2 = 0.5$, $\sigma_\alpha^2 = 0.5$, and $\sigma_\beta^2 = 0.5$.

3. Using the simulated parameters in step 1, simulate a point pattern from a NHPPP with intensity given as in equation (3.3).

4. For each spatial location simulate the Hawkes process parameters $\mu(s_l), \alpha(s_l)$, and $\beta(s_l)$, for $l = 1, \ldots, L$, using the random fields with parameters obtained in step 3.

5. Set the temporal window of observation as $\mathcal{W}_t = \mathcal{T} = [0, 20]$ (i.e., $T = 20$). For the temporal window $\mathcal{W}_t$ and the Hawkes process parameters obtained in step 4, simulate $\psi_{K(s_l)}(s_l) = \{t_1(s_l), \ldots, t_{K(s_l)}(s_l)\} \sim Hawkes(\mu(s_l), \alpha(s_l), \beta(s_l), T)$, for $l = 1, \ldots, L$.

6. The space-time point pattern is given by $\varphi_n = \{(s_l, t_k(s_l)), \ k = 1, \ldots, K(s_l), \ l = 1, \ldots, L\}$. 

126
3.5 Simulation results

In this section we present the results of the simulated example to illustrate that the algorithm works. We shall present the results in two parts; first for the spatial component and then conditioning on location we shall present the results for the temporal component. A total of 9144 events were simulated consisting of 202 spatial locations and at each of the location there are a number of tame stamps at which the events occurred. we here by give the results for the recovery of the spatial process and the temporal processes conditioned on each of the locations.

3.5.1 Recovering the Spatial (Location) Process

In Figures 3.1 and 3.2 we present the intensity function of the true process of spatial point process (202 spatial events in the point pattern), in two and three dimensions respectively. The second row presents the MAPE (Maximum a posteriori estimator) surface based on the MAP estimates of the mixture intensity parameters, which was obtained 10,000 posterior realizations. Here we used the `sppmix` R package function `est_mix_damcmc` (Chen et al. (2017), https://cran.r-project.org/web/packages/sppmix/index.html). It is evident from the trace plots (omitted) that no label switching was present. This was confirmed using the `Check()` function `sppmix` package for checking presence of label switching. Further issues of convergence of the chains was done using the running means plots (omitted) and we were satisfied that the chains indeed converged. The MAP estimates of the bivariate normal mixture distribution that is used to model the spatial variation are given the are in table 3.1. Notice that there is some relabeling of the components. All the parameters are with the 95% credible Intervals.

3.5.2 Recovering the Temporal processes conditioned on location

Now given the location of an event, we generate the temporal processes, i.e., a Hawkes process arises at each observed location. We used 10,000 iterations of the MCMC to sample
the parameters of the fields used in creating the Hawkes process. It is worth noting that the
at each of the locations the parameters governing the temporal process were different and
arose from the distributions that are detailed in equations 3.18, 3.19 and 3.20. Table 3.2 gives
the summary of the estimates of the parameters governing the Hawkes process (temporal
process) for each of the locations.

Due to large number of locations we shall on give summaries of the temporal process
(Hawkes) for only a few selected locations by indexed by 1, 20, 40, 60, 80, 100, 120, 140, 160,
180, and 202 (chosen arbitrary to demonstrate). The summaries of the estimates of Hawkes
process parameters are given in Tables 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, 3.12 and 3.13. It
worth noting that not all the hawkes process parameters are within the 95% credible sets,
however they are pretty close. To examine the behavior of the chains we first examined the
trace plots, histograms and the ACFs of the chains arising from the marginal distributions
of the parameters. (see Figures 3.3, 3.5, 3.7, 3.9, 3.11, 3.13, 3.15, 3.17, 3.19, 3.21, and 3.23).
Clearly the chains are converging (it was also established using the running means plots that
are omitted here). We also provided that the estimated (using the posterior mean and median
of the parameters at each location), the 95% credible intervals and the true hawkes intensity
functions. (see Figures 3.4, 3.6, 3.8, 3.10, 3.12, 3.14, 3.16, 3.18, 3.20, 3.22, 3.24) This could
be as a result of non-identifiability, something that can be investigated as part of future work.

3.6 Discussion

In this Chapter, we presented a hierarchical Bayesian framework for modeling the intensi-
ty function of the STPP via conditioning on the location component of the event. This type
of conditioning renders the locations fixed, which may not be appropriate from an applica-
tion point of view when looking at real life data as they arise from (3.2). Indeed, when we
condition on locations being fixed, then it is counter-intuitive to fit a model for the location
component (e.g., fixed locations for monitoring stations, hospitals or police stations), and we could only report the model fits to the remaining part of the joint distribution, i.e., the marked temporal process part \( \prod_{l=1}^{L} f(\varphi^s|s_l, \theta(s_l)) \) of (3.2). Therefore, although we may fit the location part of the joint distribution, in some scenarios it is only performed for mathematical reasons so that we can write down the joint distribution of the events and the results we are mostly interested in from an application point of view involve the marked temporal portion of the joint. The models presented are highly amenable to changes, including choice of random field, covariance function, and evolution model for the dynamically and spatially varying mark probability fields.
3.A Full Posterior Distributions

In this appendix we present the full conditional distributions for the space-time NHPPP via conditioning on location based on the full posterior distribution (3.26).

3.A.1 Updating $\lambda$

The full-conditional distribution for $\lambda$ is simply $\lambda|\cdot \sim G(\sum_{l=1}^{L} n(s_l) + a_\lambda, (1 + 1/b_\lambda)^{-1})$.

3.A.2 Updating $\lambda_0$

The full-conditional distribution for $\lambda_0$ is simply $\lambda_0|\cdot \sim G(L + a_\lambda, (1 + 1/b_\lambda)^{-1})$.

3.A.3 Updating the membership indicator vector $z_i$

The full-conditional distribution for the membership indicator vector $z_i$ is given by

$$z_i|\cdot \sim Multinomial(1, q_{l,1}, \ldots, q_{l,m}), \quad (3.27)$$

where

$$q_{l,j} = \frac{p_j \varphi_j(s_l | \mu_j, \Sigma_j)}{\sum_{r=1}^{m} p_r \varphi_l(s_l | \mu_r, \Sigma_r)}, \quad (3.28)$$

for $j = 1, \ldots, m, \ l = 1, \ldots, L$.

3.A.4 Updating the component probabilities $p$

The full-conditional distribution for the component probabilities vector $p$ is given by

$$p|\cdot \sim Dirichlet(d_1 + n_1, \ldots, d_m + n_m), \quad (3.29)$$
where \( n_j = \sum_{l=1}^{L} z_{l,j} \) for \( j = 1, \ldots, m \), represents the number of locations assigned to the \( j^{th} \) component.

### 3.A.5 Updating the \( j^{th} \) component mean \( \mu_j \)

The full-conditional distribution for the \( j^{th} \) component mean at time \( t_k \) is given by

\[
\mu_j | \cdot \sim N_2(M_j, V_j),
\]

(3.30)

where

\[
M_j = V_j (n_j \Sigma_j^{-1} \bar{s}_j + A^{-1} \bar{s}),
\]

and

\[
V_j = (n_j \Sigma_j^{-1} + A^{-1})^{-1},
\]

with \( \bar{s}_j = \frac{1}{n_j} \sum_{l=1}^{L} z_{l,j} s_l \), for \( j = 1, \ldots, m \), and \( \bar{s} = \frac{1}{L} \sum_{l=1}^{L} s_l \).

### 3.A.6 Updating the \( j^{th} \) component covariance matrix \( \Sigma_j \)

The full-conditional distribution for the \( j^{th} \) component covariance matrix is given by

\[
\Sigma_j^{-1} | \cdot \sim \text{Wishart}_2((2a_1 + n_j), Q_j)
\]

(3.31)

where

\[
Q_j = \left[ 2B + \sum_{l=1}^{L} z_{l,j}(s_l - \mu_j)(s_l - \mu_j)^T \right]^{-1},
\]

for \( j = 1, \ldots, m \).
3.A.7 Updating B

The full-conditional distribution for the \( j \)th component covariance matrix is given by

\[
B|· \sim \text{Wishart}_2((2g + 2ma_1), Q_B)
\]  
(3.32)

where

\[
Q_B = \left[ 2H + \sum_{j=1}^{m} \Sigma_j^{-1} \right]^{-1}.
\]

3.A.8 Updating the Hawkes parameters \( \mu(s) \), \( \alpha(s) \) and \( \beta(s) \)

The full-conditional distribution of the Hawkes process parameters \( \mu(s) \), \( \alpha(s) \) and \( \beta(s) \) conditional on location is given by

\[
\pi(\mu(s)|·) \propto \frac{\exp(-\Lambda(s, T))}{\mu((s))} \times \exp\left( -\frac{\left( \log(\mu(s)) - \zeta_\mu \right)^2}{2\sigma^2_\mu} \right) \times \prod_{k=1}^{n(s)} \left( \mu(s) + \alpha(s)\beta(s) \sum_{i:t_i < t_k(s)} e^{-\beta(s)(t_k(s)-t_i)} \right),
\]  
(3.33)

\[
\pi(\alpha(s)|·) \propto \frac{\exp(-\Lambda(s, T))}{\alpha((s))(1 - \alpha((s))))} \times \exp\left( -\frac{\left( \log(\frac{\alpha(s)}{1 - \alpha(s)}) - \zeta_\alpha \right)^2}{2\sigma^2_\alpha} \right) \times \prod_{k=1}^{n(s)} \left( \mu(s) + \alpha(s)\beta(s) \sum_{i:t_i < t_k(s)} e^{-\beta(s)(t_k(s)-t_i)} \right),
\]  
(3.34)

\[
\pi(\beta(s)|·) \propto \frac{\exp(-\Lambda(s, T))}{\beta((s))} \times \exp\left( -\frac{\left( \log(\beta(s)) - \zeta_\beta \right)^2}{2\sigma^2_\beta} \right) \times \prod_{k=1}^{n(s)} \left( \mu(s) + \alpha(s)\beta(s) \sum_{i:t_i < t_k(s)} e^{-\beta(s)(t_k(s)-t_i)} \right),
\]  
(3.35)
Then the full-conditional distribution of \( \mu(s) \) is given by

\[
\Lambda(s, T) = \int_0^T \left( \mu(s) + \alpha(s) \beta(s) \sum_{i:t_i<t} e^{-\beta(s)(t-t_i)} \right) dt.
\] (3.36)

Clearly it is complicated to sample from the full conditional distributions of \( 
\pi(\mu(s)|\cdot), \pi(\alpha(s)|\cdot), \) and \( \pi(\beta(s)|\cdot) \) given by

3.33, 3.34 and 3.35 therefore we use the metropolis algorithm to obtain the posterior samples of the parameters of the hawkes process. The proposal distributions for the Hawkes process parameters \( \mu(s), \alpha(s) \) and \( \beta(s) \) are given by

\( \mu(s) \sim N(\mu(s), \sigma_\mu^2) \), truncated in \((0, \infty), \alpha(s) \sim N(\alpha(s), \sigma_\alpha^2) \), truncated in \((0, 1), \) and \( \beta(s) \sim N(\beta(s), \sigma_\beta^2) \), truncated in \((0, \infty). \) The Metropolis-Hasting Ratios for updating \( \mu(s), \alpha(s) \) and \( \beta(s), \) at each location \( s \) are given by

\[
MH_{\mu(s)} = \frac{\pi(\mu(s)^{\text{prop}})|\cdot)}{\pi(\mu(s)|\cdot)} \frac{p(\mu(s))}{p(\mu(s)^{\text{prop}})},
\] (3.37)

\[
MH_\alpha = \frac{\pi(\alpha(s)^{\text{prop}})|\cdot)}{\pi(\alpha(s)|\cdot)} \frac{p(\alpha(s))}{p(\alpha(s)^{\text{prop}})},
\] (3.38)

\[
MH_{\beta(s)} = \frac{\pi(\beta(s)^{\text{prop}})|\cdot)}{\pi(\beta(s)|\cdot)} \frac{p(\beta(s))}{p(\beta(s)^{\text{prop}})},
\] (3.39)

3.A.9 Updating \( \zeta_\mu \)

Let \( \mu(s_{1:L}) = [\log(\mu(s_1)), \ldots, \log(\mu(s_t))]^T, \) and \( C_{\sigma_\mu^2} = \text{diag}(L). \) Then the full-conditional distribution of \( \zeta_\mu \) is given by

\[
\zeta_\mu | \cdot \sim N \left( \frac{\sigma_{\zeta_\mu}^2 1^T_L C_{\zeta_\mu}^{-1} \mu(s_{1:L}) + \mu_{\zeta_\mu}}{\sigma_{\zeta_\mu}^2 1^T_L C_{\zeta_\mu}^{-1} 1_L + 1}, \frac{\sigma_{\zeta_\mu}^2}{\sigma_{\zeta_\mu}^2 1^T_L C_{\zeta_\mu}^{-1} 1_L + 1} \right).
\] (3.40)

3.A.10 Updating \( \zeta_\alpha \)

Let \( \alpha(s_{1:L}) = [\log(\alpha(s_1)/(1-\alpha(s_1))), \ldots, \log(\alpha(s_t)(1-\alpha(s_t)))]^T, \) and \( C_{\sigma_\alpha^2} = \text{diag}(L). \) Then the full-conditional distribution of \( \zeta_\alpha \) is given by

\[
\zeta_\alpha | \cdot \sim N \left( \frac{\sigma_{\zeta_\alpha}^2 1^T_L C_{\zeta_\alpha}^{-1} \alpha(s_{1:L}) + \mu_{\zeta_\alpha}}{\sigma_{\zeta_\alpha}^2 1^T_L C_{\zeta_\alpha}^{-1} 1_L + 1}, \frac{\sigma_{\zeta_\alpha}^2}{\sigma_{\zeta_\alpha}^2 1^T_L C_{\zeta_\alpha}^{-1} 1_L + 1} \right).
\] (3.41)
3.A.11 Updating $\zeta_\beta$

Let $\beta(s_{1:L}) = [\log(\beta(s_1)), \ldots, \log(\beta(s_L))]^T$, and $C_{\sigma_\beta^2} = \text{diag}(L)$. Then the full-conditional distribution of $\zeta_\beta$ is given by

$$\zeta_\beta | \cdot \sim N\left(\frac{\sigma_{\zeta_\beta}^2 1_L^T C_{\zeta_\beta}^{-1} \beta(s_{1:L}) + \beta_{\zeta_\beta}}{\sigma_{\zeta_\beta}^2 1_L^T C_{\zeta_\beta}^{-1} 1_L + 1}, \frac{\sigma_{\zeta_\beta}^2}{\sigma_{\zeta_\beta}^2 1_L^T C_{\zeta_\beta}^{-1} 1_L + 1}\right).$$  \hspace{1cm} (3.42)

3.A.12 Updating $\sigma_\mu^2$

Let $\mu(s_{1:L}) = [\log(\mu(s_1)), \ldots, \log(\mu(s_L))]^T$ and $C_{\sigma_\mu^2} = \text{diag}(L)$. Then the full-conditional of $\sigma_\mu^2$ is given by

$$\sigma_\mu^2 | \cdot \sim IG\left(\frac{2a_{\sigma_\mu} + 1}{2}, \left(\frac{1}{2}(\mu(s_{1:L}) - \zeta_\mu 1_L)^T C_{\sigma_\mu}^{-1} (\mu(s_{1:L}) - \zeta_\mu 1_L) + 1\right)^{-1}\right).$$  \hspace{1cm} (3.43)

3.A.13 Updating $\sigma_\alpha^2$

Let $\alpha(s_{1:L}) = [\log(\alpha(s_1)/(1 - \alpha(s_1))), \ldots, \log(\alpha(s_L)/(1 - \alpha(s_L)))]^T$, and $C_{\sigma_\alpha^2} = \text{diag}(L)$. Then the full-conditional of $\sigma_\alpha^2$ is given by

$$\sigma_\alpha^2 | \cdot \sim IG\left(\frac{2a_{\sigma_\alpha} + 1}{2}, \left(\frac{1}{2}(\alpha(s_{1:L}) - \zeta_\alpha 1_L)^T C_{\sigma_\alpha}^{-1} (\alpha(s_{1:L}) - \zeta_\alpha 1_L) + 1\right)^{-1}\right).$$  \hspace{1cm} (3.44)

3.A.14 Updating $\sigma_\beta^2$

Let $\beta(s_{1:L}) = [\log(\beta(s_1)), \ldots, \log(\beta(s_L))]^T$, and $C_{\sigma_\beta^2} = \text{diag}(L)$. Then the full-conditional of $\sigma_\beta^2$ is given by

$$\sigma_\beta^2 | \cdot \sim IG\left(\frac{2a_{\sigma_\beta} + 1}{2}, \left(\frac{1}{2}(\beta(s_{1:L}) - \zeta_\beta 1_L)^T C_{\sigma_\beta}^{-1} (\beta(s_{1:L}) - \zeta_\beta 1_L) + 1\right)^{-1}\right).$$  \hspace{1cm} (3.45)
### 3.B Figures

<table>
<thead>
<tr>
<th>Component</th>
<th>Parameter</th>
<th>True Value</th>
<th>Posterior Mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(3)</td>
<td>$\lambda$</td>
<td>200</td>
<td>201.9252</td>
<td>[ 175.523 , 230.9596 ]</td>
</tr>
<tr>
<td></td>
<td>$p_1$</td>
<td>0.2</td>
<td>0.3022</td>
<td>[ 0.2398 , 0.3681 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>3</td>
<td>4.9995</td>
<td>[ 4.7201 , 5.2814 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>3</td>
<td>7.8814</td>
<td>[ 7.6547 , 8.0932 ]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.5</td>
<td>1.2604</td>
<td>[ 0.8871 , 1.7865 ]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>0</td>
<td>-0.102</td>
<td>[ -0.3609 , 0.1293 ]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.5</td>
<td>0.6683</td>
<td>[ 0.4542 , 0.97 ]</td>
</tr>
<tr>
<td>2(1)</td>
<td>$p_2$</td>
<td>0.3</td>
<td>0.4738</td>
<td>[ 0.4052 , 0.5435 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>5</td>
<td>6.979</td>
<td>[ 6.8249 , 7.1336 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>8</td>
<td>4.0735</td>
<td>[ 3.9219 , 4.2259 ]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>1</td>
<td>0.5698</td>
<td>[ 0.4246 , 0.768 ]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>0</td>
<td>-0.1718</td>
<td>[ -0.3128 , -0.0528 ]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>1</td>
<td>0.5396</td>
<td>[ 0.3921 , 0.749 ]</td>
</tr>
<tr>
<td>3(2)</td>
<td>$p_3$</td>
<td>0.5</td>
<td>0.2241</td>
<td>[ 0.1701 , 0.2851 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>7</td>
<td>3.1978</td>
<td>[ 3.0136 , 3.3939 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>4</td>
<td>2.9578</td>
<td>[ 2.7537 , 3.1719 ]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{11}$</td>
<td>0.5</td>
<td>0.3728</td>
<td>[ 0.2212 , 0.5998 ]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{12}$</td>
<td>-0.1</td>
<td>0.024</td>
<td>[ -0.1094 , 0.1641 ]</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{22}$</td>
<td>0.5</td>
<td>0.5036</td>
<td>[ 0.3303 , 0.7615 ]</td>
</tr>
</tbody>
</table>

Table 3.1: Summary of the true and estimated values of the mixture parameters. The true component assignments are shown in the first column, along with the component assignments in the posterior simulation (in parenthesis).
Figure 3.1: 2-d plots of the true Poison intensity surfaces (left) and the corresponding PME intensity surfaces (right). The point patterns are also shown and the gray crosses represent the mixture component means.

Figure 3.2: 3-d plots of the true Poison intensity surfaces (left) and the corresponding PME intensity surfaces (right).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\zeta_\mu$</td>
<td>0</td>
<td>0.002</td>
<td>[-0.0949 , 0.1001]</td>
</tr>
<tr>
<td>$\sigma^2_\mu$</td>
<td>0.5</td>
<td>0.4773</td>
<td>[0.2086 , 0.8746]</td>
</tr>
<tr>
<td>$\zeta_\alpha$</td>
<td>0</td>
<td>0.0015</td>
<td>[-0.1152 , 0.1215]</td>
</tr>
<tr>
<td>$\sigma^2_\alpha$</td>
<td>0.5</td>
<td>0.717</td>
<td>[0.4922 , 1.0489]</td>
</tr>
<tr>
<td>$\zeta_\beta$</td>
<td>0</td>
<td>0.0023</td>
<td>[-0.1209 , 0.1217]</td>
</tr>
<tr>
<td>$\sigma^2_\beta$</td>
<td>0.5</td>
<td>0.7639</td>
<td>[0.518 , 1.0854]</td>
</tr>
</tbody>
</table>

Table 3.2: Summary of spatial parameters
<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\lambda(s_1)$</td>
<td>116.9519</td>
<td>[ 96.9416 , 139.1417 ]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu(s_1)$</td>
<td>3.2636</td>
<td>0.7508</td>
<td>[ 0.1966 , 1.6548 ]</td>
</tr>
<tr>
<td></td>
<td>$\alpha(s_1)$</td>
<td>0.4618</td>
<td>0.6764</td>
<td>[ 0.3998 , 0.8876 ]</td>
</tr>
<tr>
<td></td>
<td>$\beta(s_1)$</td>
<td>1.1159</td>
<td>3.4063</td>
<td>[ 0.9422 , 8.9779 ]</td>
</tr>
</tbody>
</table>

Table 3.3: Summary posterior Hawkes intensity function parameters at location index=1

Figure 3.3: Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=1.
Figure 3.4: Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=1.

Table 3.4: Summary posterior Hawkes intensity function parameters at location index=20

<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>$\lambda(s_{20})$</td>
<td>37.0465</td>
<td>0.7265</td>
<td>[0.2089, 1.5331]</td>
</tr>
<tr>
<td></td>
<td>$\mu(s_{20})$</td>
<td>1.4685</td>
<td>0.5886</td>
<td>[0.4005, 0.7773]</td>
</tr>
<tr>
<td></td>
<td>$\alpha(s_{20})$</td>
<td>0.2257</td>
<td>1.7994</td>
<td>[0.7925, 3.9156]</td>
</tr>
<tr>
<td></td>
<td>$\beta(s_{20})$</td>
<td>0.4027</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.5: Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=20.
Figure 3.6: Displaying the estimated Hawkes process based on the posterior means (magenta), median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=20

<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>$\lambda(s_{40})$</td>
<td>0.6676</td>
<td>0.6816</td>
<td>[0.1855, 1.514]</td>
</tr>
<tr>
<td></td>
<td>$\mu(s_{40})$</td>
<td>0.667</td>
<td>0.6066</td>
<td>[0.4063, 0.8036]</td>
</tr>
<tr>
<td></td>
<td>$\alpha(s_{40})$</td>
<td></td>
<td>0.6066</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta(s_{40})$</td>
<td>1.2884</td>
<td>1.7867</td>
<td>[0.843, 3.7324]</td>
</tr>
</tbody>
</table>

Table 3.5: Summary posterior Hawkes intensity function parameters at location index=40
Figure 3.7: Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=40.
Figure 3.8: Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=40.

<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>$\lambda(s_{60})$</td>
<td>38.9344</td>
<td>27.6479, 52.4181</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu(s_{60})$</td>
<td>1.6097</td>
<td>0.7295, 1.6541</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha(s_{60})$</td>
<td>0.5194</td>
<td>0.4132, 0.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta(s_{60})$</td>
<td>0.9768</td>
<td>0.8205, 4.0852</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.6: Summary posterior Hawkes intensity function parameters at location index=60
Figure 3.9: Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=60.
Figure 3.10: Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=60.

<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>$\lambda(s_{80})$</td>
<td>38.9399</td>
<td>38.9399</td>
<td>[ 27.8635 , 52.158 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu(s_{80})$</td>
<td>1.0369</td>
<td>0.7626</td>
<td>[ 0.2079 , 1.7202 ]</td>
</tr>
<tr>
<td></td>
<td>$\alpha(s_{80})$</td>
<td>0.5388</td>
<td>0.5934</td>
<td>[ 0.4062 , 0.7865 ]</td>
</tr>
<tr>
<td></td>
<td>$\beta(s_{80})$</td>
<td>0.8731</td>
<td>1.7731</td>
<td>[ 0.8266 , 3.7916 ]</td>
</tr>
</tbody>
</table>

Table 3.7: Summary posterior Hawkes intensity function parameters at location index=80
Figure 3.11: Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=80.
Figure 3.12: Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=80.

<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>$\lambda(s_{100})$</td>
<td>74.0909</td>
<td>74.0909</td>
<td>[58.1458, 91.4254]</td>
</tr>
<tr>
<td></td>
<td>$\mu(s_{100})$</td>
<td>2.1063</td>
<td>0.6953</td>
<td>[0.18, 1.5623]</td>
</tr>
<tr>
<td></td>
<td>$\alpha(s_{100})$</td>
<td>0.5327</td>
<td>0.6452</td>
<td>[0.3902, 0.8486]</td>
</tr>
<tr>
<td></td>
<td>$\beta(s_{100})$</td>
<td>2.5321</td>
<td>2.7289</td>
<td>[0.9196, 5.9839]</td>
</tr>
</tbody>
</table>

Table 3.8: Summary posterior Hawkes intensity function parameters at location index=100
Figure 3.13: Trace plots, histograms and ACF's for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=100.
Figure 3.14: Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=100.

<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>$\lambda(s_{120})$</td>
<td>7.9244</td>
<td>7.9244</td>
<td>[3.4222, 14.2774]</td>
</tr>
<tr>
<td></td>
<td>$\mu(s_{120})$</td>
<td>0.4662</td>
<td>0.8881</td>
<td>[0.2189, 2.029]</td>
</tr>
<tr>
<td></td>
<td>$\alpha(s_{120})$</td>
<td>0.4764</td>
<td>0.5196</td>
<td>[0.4142, 0.6393]</td>
</tr>
<tr>
<td></td>
<td>$\beta(s_{120})$</td>
<td>0.5909</td>
<td>1.1744</td>
<td>[0.7369, 2.0047]</td>
</tr>
</tbody>
</table>

Table 3.9: Summary posterior Hawkes intensity function parameters at location index=120.
Figure 3.15: Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=120.
Figure 3.16: Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments location index=120.

<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>140</td>
<td>$\lambda(s_{140})$</td>
<td>26.9521</td>
<td>0.1766 , 1.739</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu(s_{140})$</td>
<td>0.5914</td>
<td>0.1766 , 1.739</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha(s_{140})$</td>
<td>0.4235</td>
<td>0.4068 , 0.745</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta(s_{140})$</td>
<td>1.8245</td>
<td>0.8157 , 3.4651</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.10: Summary posterior Hawkes intensity function parameters at location index=140
Figure 3.17: Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=140.
Figure 3.18: Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=140.

<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>160</td>
<td>$\lambda(s_{160})$</td>
<td>29.0536</td>
<td>29.0536</td>
<td>[ 19.4115, 40.4495 ]</td>
</tr>
<tr>
<td></td>
<td>$\mu(s_{160})$</td>
<td>0.7411</td>
<td>0.7361</td>
<td>[ 0.1889, 1.6261 ]</td>
</tr>
<tr>
<td></td>
<td>$\alpha(s_{160})$</td>
<td>0.5609</td>
<td>0.572</td>
<td>[ 0.4047, 0.7424 ]</td>
</tr>
<tr>
<td></td>
<td>$\beta(s_{160})$</td>
<td>0.6113</td>
<td>1.6785</td>
<td>[ 0.793, 3.4377 ]</td>
</tr>
</tbody>
</table>

Table 3.11: Summary posterior Hawkes intensity function parameters at location index=160.
Figure 3.19: Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=160.
Figure 3.20: Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=160.

<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>180</td>
<td>$\lambda(s_{180})$</td>
<td>10.0967</td>
<td>10.0967</td>
<td>[4.8866, 17.3021]</td>
</tr>
<tr>
<td></td>
<td>$\mu(s_{180})$</td>
<td>0.3337</td>
<td>0.8658</td>
<td>[0.2177, 1.8998]</td>
</tr>
<tr>
<td></td>
<td>$\alpha(s_{180})$</td>
<td>0.4337</td>
<td>0.5223</td>
<td>[0.4048, 0.6607]</td>
</tr>
<tr>
<td></td>
<td>$\beta(s_{180})$</td>
<td>1.2663</td>
<td>1.2283</td>
<td>[0.7425, 2.228]</td>
</tr>
</tbody>
</table>

Table 3.12: Summary posterior Hawkes intensity function parameters at location index=180.
Figure 3.21: Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=180.
Figure 3.22: Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=180.

Table 3.13: Summary posterior Hawkes intensity function parameters at location index=202

<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Credible set</th>
</tr>
</thead>
<tbody>
<tr>
<td>202</td>
<td>( \lambda(s_{202}) )</td>
<td>1.0528</td>
<td>60.9237</td>
<td>[46.5021, 77.5697]</td>
</tr>
<tr>
<td></td>
<td>( \mu(s_{202}) )</td>
<td>0.5336</td>
<td>0.72</td>
<td>[0.1922, 1.616]</td>
</tr>
<tr>
<td></td>
<td>( \alpha(s_{202}) )</td>
<td>1.3062</td>
<td>2.1824</td>
<td>[0.8456, 4.7961]</td>
</tr>
<tr>
<td></td>
<td>( \beta(s_{202}) )</td>
<td>0.388</td>
<td>1.616</td>
<td>[0.388, 0.8308]</td>
</tr>
</tbody>
</table>

location 202
Figure 3.23: Trace plots, histograms and ACF’s for $\mu(s)$, $\alpha(s)$ and $\beta(s)$. The red dashed line indicates the posterior means for the model parameters, $\mu(s)$, $\alpha(s)$ and $\beta(s)$ at location index=202.
Figure 3.24: Displaying the estimated Hawkes process based on the posterior means (magenta), posterior median (cyan), 95% credible intervals (lower-black, upper-blue) of the model parameters. The temporal point pattern is shown at the bottom of the figure using red line segments at location index=202.
Chapter 4

Markov Marked Spatial-Temporal Point Process Models

4.1 Introduction

As discussed in Chapter 1, an additional assumption that distinguishes point patterns is whether or not the locations of the events are independent from each other or if they interact in a systematic way. Under independence, the Poisson point process model is a natural starting point as we have illustrated in Chapters 2 and 3, whereas, for models with interactions, a Markov (or Gibbs) point process modeling approach is the usual way to proceed. For Poisson models, the choice of intensity function leads to different modeling approaches, while for Gibbs models, the Papangelou conditional intensity is the focus (Van Lieshout, 2000). Moreover, the type of statistical framework employed, classical or Bayesian, is another major difference between the types of models we can propose for point patterns. In this chapter we study Markov marked space-time point processes (MMSTPPs) in a fully hierarchical Bayesian framework.

Current contributions to the literature on (marked) space-time point processes either assume that the model parameters are the same across time or that the parameters depend on time but are not dynamic. For example, even though a model parameter \( \theta_t \) may depend on the time \( t \), there is no mechanism that evolves \( \theta_t \) into \( \theta_{t+1} \). A recent contribution related to our work is King et al. (2012), where the authors illustrated (non-dynamic) Gibbs space-time models with mixed effects in order to model interacting locations of muskoxen herds.
The authors utilized the “area interaction process,” a specific type of Gibbs model, along with elevation as covariate information, but their model was not dynamic so that they could not perform forecasting. In addition, their model required the assumption of conditional independence of locations across time given the covariate in order to be able to approximate the likelihood of the space-time point pattern.

The aforementioned issues with Kottas et al. (2012) illustrate some of the difficulties one faces when working with such models, and as we discuss below, the methodology considered here mitigates these difficulties. In particular, this is where hierarchical Bayesian models are invaluable, since they allow us to introduce dynamic process models that make prediction of future states of the point process more accurate and efficient. Introducing dynamic process models for the parameters of the point process is one of the major contributions of this chapter. We begin by describing the general components of the space-time models, including three approaches to modeling the main effect and a common approach to modeling the interaction effect.

4.2 Markov marked space-time point processes

Consider $N$ individuals and suppose that we observe their locations at $T$ different time periods; namely, we have a point pattern $\varphi_n = \{x_1, \ldots, x_n\}$, $n = NT$, fixed, with the $i^{th}$ point $x_i = (s_i, t_i, z_i) \in X = W_s \times W_t \times W_z$, (where $W_s \subset \mathbb{R}^d$, $W_t = \{1, \ldots, T\}$, $W_z = \{1, \ldots, N\}$, with $|X| = n|W_z|$), containing three types of information as follows:

1) $s_i$ is the location of the $i^{th}$ point in space, $s_i \in W_s$,
2) $t_i$ denotes the time at which this point was observed, $t_i \in W_t$, and
3) $z_i$ denotes a mark, for example, to which individual this point corresponds to; that is, $z_i \in W_z$. In general, we consider $M$ discrete marks in a mark space $\mathcal{M} = \{m_1, \ldots, m_M\}$.

In order to model the point pattern $\varphi_n$ we turn to Markov marked point processes in order to be able to write down the likelihood. In particular, first assume that we introduce a neighborhood structure via a symmetric and reflexive relation $\sim$ on $X$, so that two points
x and y are said to be neighbours if and only if $x \sim y$. Following Diggle (2006, pg 113), we can write the joint distribution of the point pattern $\varphi_n$ using

$$f_n(\varphi_n) = \frac{1}{c_n} \exp \left\{ \sum_{i=1}^{n} g_1(x_i) + \sum_{1 \leq i \neq j \leq n} g_{1,2}(x_i, x_j) + \ldots + g_{1,2,\ldots,n}(x_1, \ldots, x_n) \right\},$$  \hspace{1cm} (4.1)$$

for appropriate interaction functions $g$ that vanish if their arguments are not in a clique. The choice of relation is critical in this case, and will lead to different types of models that may include behaviours such as attraction or repulsion of the individuals, provided that certain conditions are satisfied. We write $E(\varphi_n)$ to denote the “energy” function

$$E(\varphi_n) = \sum_{i=1}^{n} g_1(x_i) + \sum_{1 \leq i \neq j \leq n} g_{1,2}(x_i, x_j) + \ldots + g_{1,2,\ldots,n}(x_1, \ldots, x_n).$$

### 4.3 The non-homogeneous pairwise MMSTPP model

We will consider only first and second order $g$-functions in this development; that is, we assume that

$$f_n(\varphi_n) = \frac{1}{c_n} \exp \left\{ \sum_{i=1}^{n} g_1(x_i) + \sum_{1 \leq i \neq j \leq n} g_{1,2}(x_i, x_j) \right\}.$$  \hspace{1cm} (4.2)$$

The resulting model is known as a non-homogeneous pairwise MMSTPP. It is straightforward to see that the relation $\sim$ defined by $x_i \sim x_j \leftrightarrow \|s_i - s_j\| \leq \rho$, and $(t_j = t_i + 1$ (future) or $t_j = t_i$ (present) or $t_j = t_i - 1$ (past)), is symmetric and reflexive, while it does not depend on the particular individual tags, thus allowing interaction between any of the individuals, but only for locations that are closer than $\rho$ and times that are within one time frame from each other, namely, $|t_j - t_i| \leq 1$. In any other case, the two points are not neighbors and therefore we should not get any contribution in the likelihood.

#### 4.3.1 Modeling of the main effects

For the first order cliques, we choose a function $g_1$ that may depend on covariate information and describes a dependence structure (if it is desired) between location and time. In
general, we want \( g_1 \) to correspond to an overall propagation effect that depends on time and possible covariate information. For example, Illian et al. (2008) choose \( g_1(s_i) = (\theta + \phi_t)^T 1 \), to be a fixed effect plus a random effect that depends on time in their Gibbs process, so that they end up with a Strauss pairwise interaction model once they model the second order clique functions.

In our formulation, assume that we are given a covariate field \( W = \{W(s)\}_{s \in W} \); for example,

\[
W(s)_{2 \times 8} = \begin{bmatrix}
W_1(s)_{1 \times 4} & 0_{1 \times 4} \\
0_{1 \times 4} & W_2(s)_{1 \times 4}
\end{bmatrix}_{2 \times 8},
\]

with

\[
W_1(s)_{1 \times 4} = W_2(s)_{1 \times 4} = [1, s_x, s_y, s_x s_y],
\]

for \( s = [s_x, s_y] \), and

\[
\beta = \begin{bmatrix}
\beta_x \\
\beta_y
\end{bmatrix}_{8 \times 1}.
\]

In general, we choose

\[
g_1(x_i) = \log(f_Z(z_i)) + \log(\xi) - \frac{1}{2} \left( s_i - \left[ (W_1(s_i) \beta_x)^T, (W_2(s_i) \beta_y)^T \right] - \phi_t_i \right)^T \\
\times \Sigma^{-1}(t_i) \left( s_i - \left[ (W_1(s_i) \beta_x)^T, (W_2(s_i) \beta_y)^T \right] - \phi_t_i \right)
\]

\[
= \log(f_Z(z_i)) + \log(\xi) - \frac{1}{2} \left( s_i - W(s_i) \left[ \beta_x^T, \beta_y^T \right]^T - \phi_t_i \right)^T \\
\times \Sigma^{-1}(t_i) \left( s_i - W(s_i) \left[ \beta_x^T, \beta_y^T \right]^T - \phi_t_i \right),
\]

or

\[
g_1(x_i) = \log(f_Z(z_i)) + \log(\xi) - \frac{1}{2} (s_i - W(s_i) \beta - \phi_t_i)^T \Sigma^{-1}(t_i)(s_i - W(s_i) \beta - \phi_t_i), \quad (4.3)
\]

with \( f_Z(z_i) \) the mark distribution, \( \xi > 0 \), controlling the number of points in the point pattern, \( \Sigma(t_i) \) is a \( 2 \times 2 \) covariance matrix, two fixed effect \( p \times 1 \) vectors \( \beta_x \) and \( \beta_y \) that
control the effect of the covariate information $W(s)_{2 \times 2p}$ on each of the coordinates of the event location $s_i$, and a random effect $\phi_{t_i}$ that provides an overall propagation effect over time. Note that $p = 4$ in the example above. Furthermore, $g_1(x_i)$ can be taken to be independent of the individual component $z_i$ (the mark) of $x_i$, implying a discrete uniform distribution for the marks over $W = \{1, ..., N\}$. For fixed marks we do not require the term $\log(f_Z(z_i))$ in (4.3) and consequently, in all the development that follows we set $f_Z(z_i) = 1$ in the fixed mark case.

**Model 1: Linear random effect $\phi_{t_i}$**

In model 1, we consider a linear random effect in time given by $\phi_{t_i} = [a_1 t_i + b_1, a_2 t_i + b_2]^T = t_i \mathbf{a} + \mathbf{b}$, $t_i \in \mathcal{W}_t = \{1, ..., T\}$, with the vectors $\mathbf{a} = [a_1, a_2]^T$, and $\mathbf{b} = [b_1, b_2]^T$, having some distribution, e.g., $\mathbf{a} \sim \mathcal{N}_2(\mu_a, \sigma^2_a \mathbf{I}_2)$, and $\mathbf{b} \sim \mathcal{N}_2(\mu_b, \sigma^2_b \mathbf{I}_2)$. Furthermore, we choose $\Sigma(t_i) = \sigma^2 \mathbf{I}_2$, for all $i = 1, 2, ..., n$. Thus, the main effect becomes

$$g_1(x_i) = \log(f_Z(z_i)) + \log(\xi) - \frac{1}{2\sigma^2} \|s_i - W(s_i)\beta - t_i \mathbf{a} + \mathbf{b}\|_2^2. \quad (4.4)$$

**Model 2: Dynamic evolution of the random effect $\phi_{t_i}$**

For model 2, we consider a VAR(1) structure for the random effects $\phi_1, ..., \phi_T$ as follows. Assume that $\phi_0 \sim \mathcal{N}_2(\mu_0, \Sigma_0)$, and

$$\phi_t|\phi_{t-1}, \Phi, \Sigma_\phi \sim \mathcal{N}_2(\Phi \phi_{t-1}, \Sigma_\phi), \quad (4.5)$$

for $t = 1, ..., T$, for some given $\mu_0$ and $\Sigma_0$, transition matrix $\Phi$, and error covariance matrix $\Sigma_\phi$. Instead of a diagonal matrix, we choose $\Sigma(t) = \Sigma$, again non-evolving with time, for all $t = 1, 2, ..., T$. Therefore, the main effect in this formulation is given by

$$g_1(x_i) = \log(f_Z(z_i)) + \log(\xi) - \frac{1}{2}(s_i - W(s_i)\beta - \phi_{t_i})^T \Sigma^{-1}(s_i - W(s_i)\beta - \phi_{t_i}). \quad (4.6)$$
Model 3: Dynamic evolution of the random effect $\phi_{t_i}$ and the covariance matrix $\Sigma(t_i)$

For model 3, we consider again a VAR(1) structure for the random effect random effect as in model 2, so that the $\phi_t$ evolve over time dynamically according to (4.5). In addition, we assume that the variance-covariance matrices $\Sigma(t)$ depend on time and they are dynamically evolving (see Section 4.3.4 for the details). The form of the main effect in this case is given by

$$g_1(x_i) = \log(f_Z(z_i)) + \log(\xi) - \frac{1}{2}(s_i - W(s_i)\beta - \phi_t)^T \Sigma(t)^{-1}(t_i)(s_i - W(s_i)\beta - \phi_t).$$ (4.7)

### 4.3.2 Modeling of the interaction effect

The second order cliques are described using the functions $g_{1,2}$ and will be taken to be free from covariate information. Following the relation induced neighborhood structure described in beginning of this section, we choose a pairwise interaction function as follows:

$$g_{1,2}(x_i, x_j) = \begin{cases} 
0, & \text{if } \|s_i - s_j\| > \rho \text{ (too far away, marks have no effect)}, \\
0, & \text{if } \|s_i - s_j\| \leq \rho \text{ (attraction) and } |t_i - t_j| > 1, \\
\log \gamma, & \|s_i - s_j\| \leq \rho \text{ (attraction) and } |t_i - t_j| \leq 1
\end{cases}$$ (4.8)

### 4.3.3 Model distributions

Under the aforementioned assumptions for the $g$-functions, the model distribution (4.2)
assumes the form

$$f_n(\varphi_n|\theta) = \frac{1}{c_n(\theta)} \xi^n \gamma Y_n(\varphi_n; \rho) e^{W_n(\varphi_n; \theta_1)} \prod_{i=1}^{n} f_Z(z_i),$$ (4.9)
where $\theta = (\rho, \gamma, \xi, \beta, \theta_1)$ contains the model parameters, $c_n(\theta)$ is the intractable normalizing constant for the Markov distribution with $n$ events, and

$$W_n(\varphi_n; \theta_1) = -\frac{1}{2} \sum_{i=1}^{n} (s_i - W(s_i) \beta - \phi_{t_i})^T \Sigma^{-1}(t_i)(s_i - W(s_i) \beta - \phi_{t_i}), \quad (4.10)$$

represents the large scale effects and the total contribution of the covariates, with $\theta_1 = (\beta, \{\phi_{t_i}\}_{i=1}^{n}, \{\Sigma(t_i)\}_{i=1}^{n})$, and

$$Y_n(\varphi_n; \rho) = \sum_{1 \leq i \neq j \leq n} I(\|s_i - s_j\| \leq \rho) I(|t_i - t_j| \leq 1), \quad (4.11)$$

is the total number of neighbor pairs across time and space. The specific models for the main effect are obtained by appropriate choice of $\phi_{t_i}$ and $\Sigma(t_i)$ in the form of $W_n(\varphi_n; \theta_1)$. If $\gamma = 0$ the density $f_n(\varphi_n; \rho)$ becomes

$$f_n(\varphi_n) = \begin{cases} \frac{1}{c_n(\theta)} \xi^n e^{W_n(\varphi_n; \theta_1)} \prod_{i=1}^{n} f_Z(z_i), & \text{if } Y_n(\varphi_n; \rho) = 0, \\ 0, & \text{if } Y_n(\varphi_n; \rho) > 0, \end{cases} \quad (4.12)$$

so that we obtain the classic hard-core process model. If $\gamma = 1$, there is no second order clique effect and the model reduces to a NHPPP on $\mathcal{X}$. If $0 < \gamma < 1$, the process exhibits “repulsion” or “inhibition” between the events, since $Y_n(\varphi_n; \rho)$ tends to be smaller than under the Poisson model. The case $\gamma > 1$, with $g_1(\mathbf{x}_i) = \log \beta$, for some constant $\beta > 0$, was proposed by Strauss (1975) (in the unmarked case) as a model of clustering or attraction, but $(4.9)$ is not integrable for $\gamma > 1$ (Ripley, 1976). In our formulation, the form of $W_n(\varphi_n; \theta_1) < 0$ guarantees integrability even in the case where $\gamma > 1$, since in essence we work with multivariate normal integrals.

When the marks are not fixed, they require a distribution of their own. For the mark distribution we choose a discrete distribution over the mark space $\mathcal{M} = \{m_1, ..., m_M\}$; that is,

$$f_Z(m) = P(Z = m) = p_m,$$
\( m = 1, 2, \ldots, M, \sum_{i=1}^{M} p_i = 1 \). For the \( N \)-individuals model we have \( \mathcal{M} = \{1, \ldots, N\} = \mathcal{W}_s \), i.e., \( M = N \) and \( m_i = i, i = 1, 2, \ldots, N \). The latter construction introduces an additional parameter vector \( \mathbf{p} = [p_1, \ldots, p_M] \), denoting the mark probabilities, so that, \( \theta = (\rho, \gamma, \xi, \mathbf{p}, \beta, \theta_1) \) denotes all the parameters of the model at the model stage.

### 4.3.4 Process models and prior distributions

In this section we outline the prior distributions for each model, along with the dynamically evolving parameters and their models. Note that all models contain the non-dynamic parameters \( \rho, \gamma, \xi, \mathbf{p}, \) and \( \beta \), which will be modeled as follows. Assume that

\[
\gamma|a_\gamma, b_\gamma \sim G(a_\gamma, b_\gamma), \quad (4.13)
\]

\[
\xi|a_\xi, b_\xi \sim G(a_\xi, b_\xi), \quad (4.14)
\]

where \( a_\gamma, b_\gamma, a_\xi, \) and \( b_\xi \), are hyper-parameters, which are typically fixed and used to tune the samplers. Moreover, we let

\[
\rho|\rho_{\text{max}} \sim Unif(0, \rho_{\text{max}}), \quad (4.15)
\]

where \( \rho_{\text{max}} \) is chosen based on the window, e.g., if \( \mathcal{W}_s = [0, w]^2 \), then take \( \rho_{\text{max}} = \sqrt{2}w \).

When the marks are random, we also choose a Dirichlet prior for \( \mathbf{p} = [p_1, \ldots, p_M] \) for all models; that is,

\[
\mathbf{p}|\mathbf{d} \sim Dirichlet(\mathbf{d}), \quad (4.16)
\]

with \( \mathbf{d} \) typically chosen to give equal weights \textit{a priori} to each mark, i.e., \( \mathbf{d} = [1, \ldots, 1]^T \).

Finally, since the coefficients \( \beta \) of the term with the covariate contribution in \( W_n(\varphi_n; \theta_1) \) appear in all models, we set

\[
\beta|\mu_\beta, \sigma_\beta^2 \sim \mathcal{N}_{2p}(\mu_\beta, \sigma_\beta^2 \mathbf{I}_{2p}), \quad (4.17)
\]

where \( \mu_\beta, \sigma_\beta^2 \) are fixed hyper-parameters.
Model 1 priors: non-dynamic parameters

The parameters for model 1 include $\theta = (\rho, \gamma, \xi, p, \beta, \theta_1)$, with $\theta_1 = (\sigma^2, a, b)$, and are assumed to be independent of time so that the resulting MMSTPP model has parameters that are not evolving with time. Note that they generally depend on the intractable normalizing constant $c_n(\theta)$ so they will never be conjugate. In particular, we take

$$
\sigma^2 | a_\sigma, b_\sigma \sim IG(a_\sigma, b_\sigma),
$$
(4.18)

$a_\sigma > 1$, with $E(\sigma^2) = \frac{1}{b_\sigma(a_\sigma - 1)}$, and

$$
a | \mu_a, \sigma_a^2 \sim N_2(\mu_a, \sigma_a^2 I_2),
$$
(4.19)

$$
b | \mu_b, \sigma_b^2 \sim N_2(\mu_b, \sigma_b^2 I_2),
$$
(4.20)

where $a_\sigma, b_\sigma, \mu_a, \sigma_a^2, \mu_b, \sigma_b^2$ are fixed hyper-parameters. In particular, the fixed hyper-parameters were chosen in order to make the priors have the least influence on the posterior.

Model 2 priors and dynamic evolution of $\phi_{t_i}$

The parameters under model 2 are $\theta = (\rho, \gamma, \xi, p, \beta, \theta_1)$, with $\theta_1 = (\Sigma, \{\phi_{t_i}\}_{i=1}^n)$, where $\beta$ modeled using (4.17),

$$
\Sigma | \nu_\Sigma, C_\Sigma \sim Wishart_2(\nu_\Sigma, (\nu_\Sigma C_\Sigma)^{-1}),
$$
(4.21)

for some fixed hyper-parameters $\nu_\Sigma > 1$, and $C_\Sigma$, and the dynamic evolution model for $\phi_{t_i} | \phi_{t-1}, \Phi, \Sigma_\phi$ given by (4.5), with the process parameters modeled by hyper-priors

$$
\Sigma_\phi | \nu_\phi, C_\phi \sim Wishart_2(\nu_\phi, (\nu_\phi C_\phi)^{-1}),
$$
(4.22)

and

$$
vec(\Phi) | \mu_\Phi, \Sigma_\Phi \sim N_p(\mu_\Phi, \Sigma_\Phi),
$$
(4.23)
for some fixed hyper-parameters $\nu_\phi > 1$, $C_\phi$, $\mu_\phi$, and $\Sigma_\phi$.

**Model 3 priors and dynamic evolution of $\phi_{t_i}$ and $\Sigma(t_i)$**

The parameters in the setup of model 3 include $\theta = (\rho, \gamma, \xi, \eta, \beta, \theta_1)$, with $\theta_1 = \{\phi_{t_i}\}_{i=1}^n$, $\{\Sigma(t_i)\}_{i=1}^n$. The dynamic evolution model for $\phi_t|\phi_{t-1}, \Phi, \Sigma_\phi$ follows the same setup as in model 2, but now we also model $\Sigma(t)$ dynamically. An appeal to the LDU decomposition method (see p 2.36) leads to

$$
\Sigma(t) = K(t)\Gamma(t)\Gamma(t)^T K(t),
$$

where $K(t) = \text{diag}(\kappa_1(t), \kappa_2(t))$ and $\Gamma_t = \begin{pmatrix} 1 & 0 \\ \zeta(t) & 1 \end{pmatrix}$. Therefore, instead of evolving the variance-covariance $\Sigma(t)$, we evolve $\kappa_1(t)$, $\kappa_2(t)$, and $\zeta(t)$ individually, in a similar fashion as we have already illustrated in Chapter 2. In particular, since $\Sigma(t)$ is positive definite, the values of $\kappa_1(t)$ and $\kappa_2(t)$ have to be positive, so we consider AR(1) processes for $\log(\kappa_1(t))$, $\log(\kappa_2(t))$ and $\zeta(t)$. That is, we dynamically evolve the parameters of $\Sigma(t)$ using

$$
\log(\kappa_1(0))|\tilde{\mu}_{\kappa_1}, \tilde{\sigma}_{\kappa_1}^2 \sim \mathcal{N}(\tilde{\mu}_{\kappa_1}, \tilde{\sigma}_{\kappa_1}^2),
$$

$$
\log(\kappa_2(0))|\tilde{\mu}_{\kappa_2}, \tilde{\sigma}_{\kappa_2}^2 \sim \mathcal{N}(\tilde{\mu}_{\kappa_2}, \tilde{\sigma}_{\kappa_2}^2),
$$

$$
\zeta(0)|\tilde{\mu}_\zeta, \tilde{\sigma}_\zeta^2 \sim \mathcal{N}(\tilde{\mu}_\zeta, \tilde{\sigma}_\zeta^2),
$$

and for $t = 1, \ldots, T$, we set

$$
\log(\kappa_1(t)) = \varphi_{\kappa_1} \log(\kappa_1(t - 1)) + \epsilon_1,
$$

$$
\log(\kappa_2(t)) = \varphi_{\kappa_2} \log(\kappa_2(t - 1)) + \epsilon_2,
$$

and

$$
\zeta(t) = \varphi_\zeta \zeta(t - 1) + \epsilon_3.
$$
where \( \epsilon_1 \sim N(0, \sigma_{\kappa_1}^2) \), \( \epsilon_2 \sim N(0, \sigma_{\kappa_2}^2) \), and \( \epsilon_3 \sim N(0, \sigma_\zeta^2) \). Therefore, the models become

\[
\log(\kappa_1(t)|\kappa_1(t-1), \varphi_{\kappa_1}, \sigma_{\kappa_1}^2 \sim N(\varphi_{\kappa_1} \log(\kappa_1(t-1)), \sigma_{\kappa_1}^2), \tag{4.27}
\]

\[
\log(\kappa_2(t)|\kappa_2(t-1), \varphi_{\kappa_2}, \sigma_{\kappa_2}^2 \sim N(\varphi_{\kappa_2} \log(\kappa_2(t-1)), \sigma_{\kappa_2}^2), \tag{4.28}
\]

and

\[
\zeta(t)|\zeta(t-1), \varphi_\zeta, \sigma_\zeta^2 \sim N(\varphi_\zeta \zeta(t-1)), \sigma_\zeta^2), \tag{4.29}
\]

t = 1, ..., T.

To complete the hierarchical model, we assume hyper-priors \( \sigma_\zeta^2 \sim IG(a_\zeta, b_\zeta) \), \( \sigma_{\kappa_1}^2 \sim IG(a_{\kappa_1}, b_{\kappa_1}) \), and \( \sigma_{\kappa_2}^2 \sim IG(a_{\kappa_2}, b_{\kappa_2}) \), for some positive real numbers \( a_\zeta, b_\zeta, a_{\kappa_1}, b_{\kappa_1}, a_{\kappa_2} \) and \( b_{\kappa_2} \), fixed hyper-parameters were chosen in order to make the priors have the least influence on the posterior. Finally, we assume uniform hyper-prior distributions on the transition coefficients \( \varphi_\zeta, \varphi_{\kappa_1} \) and \( \varphi_{\kappa_2} \), i.e., \( \varphi_\zeta \sim U(c_\zeta, d_\zeta) \), \( \varphi_{\kappa_1} \sim U(c_{\kappa_1}, d_{\kappa_1}) \), and \( \varphi_{\kappa_2} \sim U(c_{\kappa_2}, d_{\kappa_2}) \), for some fixed hyper-parameters \( c_\zeta, d_\zeta, c_{\kappa_1}, d_{\kappa_1}, c_{\kappa_2} \) and \( d_{\kappa_2} \). In order to implement random-walk dynamic evolution we may take \( \varphi_\zeta = \varphi_{\kappa_1} = \varphi_{\kappa_2} = 1 \).

**4.3.5 Posterior sampling via double Metropolis-Hastings**

Letting \( \eta \) denote all the hyper-parameters of the prior distributions, the full posterior distribution is given by

\[
\pi(\theta|\varphi_n) \propto f_n(\varphi_n|\theta) \pi(\theta|\eta) = \frac{1}{c_n(\theta)} \xi^n \gamma Y_n(\varphi_n; \rho) e^{W_n(\varphi_n; \theta_1)} \left( \prod_{i=1}^n p_{zi} \right) \pi(\theta|\eta),
\]

and depending on the choice of \( \phi_{t_i} \) and \( \Sigma(t_i) \) in \( W_n(\varphi_n; \theta_1) \) of equation (4.10), the forms of the full posterior distributions are given in equations (4.38), (4.53) and (4.62).

There are several approaches that can be used to sample from \( \pi(\theta|\varphi_n) \) in the presence of the intractable normalizing constant \( c_n(\theta) \). A straightforward approach is to use samples from the MMSTPP and approximate the constant using Monte Carlo integration. However,
since this approximation needs to be conducted for each parameter in the Gibbs sampler, it is often computationally expensive as the dimension of $\theta$ increases. The classical approach is due to Besag (1974), who proposed an estimation method using pseudo-likelihood, which does not work well when there is strong interaction between the events. An alternative method was proposed by Geyer and Thompson (1992) utilizing importance sampling to estimate the normalizing constant, but this method only works if the proposed parameter values used in the importance function are close to the maximum-likelihood estimate of the parameters. An overview of other estimation methods can be found in the work of Mohler et al. (2011).

Herein we utilize the double Metropolis–Hastings (MH) algorithm by Liang (2010). This is an approximate version of the auxiliary variable MH algorithm (Moller et al. 2006; Murray et al. 2012). In particular, assume that we want to sample from the posterior distribution

$$
\pi(\theta|\varphi_n) \propto f_n(\varphi_n|\theta)\pi(\theta|\eta) \propto \frac{1}{c_n(\theta)}e^{E(\varphi_n;\theta)}\pi(\theta|\eta),
$$

where

$$
E(\varphi_n;\theta) = \sum_{i=1}^{n} \log(f_Z(z_i)) + n \log(\xi) + W_n(\varphi_n;\theta_1) + Y_n(\varphi_n;\rho) \log \gamma,
$$

denotes the energy function under the proposed model of equation (4.9), with $W_n(\varphi_n;\theta_1)$ given by (4.10) and $Y_n(\varphi_n;\rho)$ given by (4.11). The general double M-H algorithm is as follows.

**Algorithm: Double M-H posterior sampler for MMSTPPs**

**Step 1:** Start with an initial state $\theta^{(0)}$. For all $k \geq 0$, assume that at time $k$ the Markov chain is at $\theta^{(k)}$, and generate a proposed parameter vector $\theta^{(\text{prop})} \sim K(.|\theta^{(k)}, \varphi_n)$, from some kernel (proposal) distribution $K(.|\theta^{(k)})$.

**Step 2:** Generate an auxiliary variable (point pattern) $\psi_n \sim f_n(\psi_n|\theta^{(\text{prop})})$ using the Birth-Death algorithm (see below) with change moves only, i.e., $\psi_n$ has the same number of points as $\varphi_n$. This simulated pattern requires the first M-H step.
Step 3: The double M-H ratio is given by

$$ r(\theta^{(\text{prop})}, \theta^{(k)}, \psi_n | \varphi_n) = \frac{\pi(\theta^{(\text{prop})} | \eta) K(\theta^{(k)} | \theta^{(\text{prop})}, \varphi_n) f_n(\psi_n | \theta^{(k)}) f_n(\varphi_n | \theta^{(\text{prop})})}{\pi(\theta^{(k)} | \eta) K(\theta^{(\text{prop})} | \theta^{(k)}, \varphi_n) f_n(\varphi_n | \theta^{(k)}) f_n(\psi_n | \theta^{(\text{prop})})}, \quad (4.30) $$

and it does not depend on the normalizing constant $c_n(\theta)$. This second M-H step leads us to aptly call the method a double M-H sampler. Choosing $K(\theta^{(\text{prop})} | \theta^{(k)}, \varphi_n) = \pi(\theta^{(\text{prop})} | \eta)$, the prior distribution simplifies the M-H ratio to

$$ r(\theta^{(\text{prop})}, \theta^{(k)}, \psi_n | \varphi_n) = \exp \left\{ E(\psi_n | \theta^{(k)}) - E(\varphi_n | \theta^{(k)}) + E(\varphi_n | \theta^{(\text{prop})}) - E(\psi_n | \theta^{(\text{prop})}) \right\}. \quad (4.31) $$

Step 4: Accept the proposed $\theta^{(\text{prop})}$ with probability $\min(1, r(\theta^{(\text{prop})}, \theta^{(k)}, \psi_n | \varphi_n))$, and set $\theta^{(k+1)} = \theta^{(\text{prop})}$. If the proposed value is not accepted, remain at the previous state, i.e., $\theta^{(k+1)} = \theta^{(k)}$. Go to step 1.

4.4 Simulating point patterns from the proposed MM-STPP model

To simulate from the proposed MMSTPP we follow the standard Birth-Death scheme in this context (see below), utilizing the Papangelou conditional intensity, either with fixed $n = NT$ (see for example, (see Van Lieshout, 2000, Theorem 3.2, pg. 84) or random $n = \sum_{1\leq j \leq T} N(j)$, over a fixed window $\mathcal{X}$, where $N(j)$ denotes the number of points with time stamp $t_j$. The Papangelou conditional intensity at the event $x_0 = (s_0, t_0, z_0) \notin \varphi_n$, is given by

$$ \lambda^*(x_0 | \varphi_n) = \frac{f_n(\varphi_n \cup \{x_0\})}{f_n(\varphi_n)} = p z_0 \xi \gamma \sum_{1 \leq n} I(\|s_0 - s_j\| \leq \rho) I(|t_0 - t_j| \leq 1) \exp \left\{ -\frac{1}{2} (s_0 - W(s_0) \beta - \phi_{t_0})^T \Sigma^{-1}(t_0)(s_0 - W(s_0) \beta - \phi_{t_0}) \right\}. $$
Suppose that at time $k$ the process is at state $\varphi_{nk}^k$. The standard Birth-Death M-H algorithm offers two options for the proposal, either a birth or a death, with probabilities, say, $b(\varphi_{nk}^k)$ and $d(\varphi_{nk}^k) = 1 - b(\varphi_{nk}^k)$, respectively. If a birth move is chosen, the new point $x \in \mathcal{W}$, is sampled from some birth (proposal) distribution with density $B(\varphi_{nk}^k, \cdot)$ with respect to Lebesgue or counting measure. In the case where a death move is chosen, one of the points $x$ in $\varphi_{nk}^k$ is randomly chosen according to some death distribution $D(\varphi_{nk}^k, \cdot)$.

We assume that births and deaths are equally likely, so that

$$b(\varphi_n) = 1/2 = d(\varphi_n),$$

and assume that the proposed births are sampled uniformly from

$$B(\varphi_n, x) = \frac{1}{|\mathcal{X}|} = \frac{1}{n|\mathcal{W}|},$$

and the proposed point for deletion is selected uniformly, that is,

$$D(\varphi_n, x) = 1/\text{card}(\varphi_n) = 1/n,$$

for any $x \in \mathcal{X}$. Then the Birth-Death M-H ratio reduces to

$$r(\varphi_n, x) = \frac{|\mathcal{X}|}{\text{card}(\varphi_n) + 1} \frac{f_n(\varphi_n \cup \{x\})}{f_n(\varphi_n)} = \frac{n|\mathcal{W}|}{n + 1} \lambda^*(x|\varphi_n),$$

so that the birth acceptance probability is given by

$$A(\varphi_{nk}^k, \varphi_{nk}^k \cup \{x\}) = \min \left\{ 1, r(\varphi_{nk}^k, x) \right\} = \min \left\{ 1, \frac{n|\mathcal{W}|}{n + 1} \lambda^*(x|\varphi_{nk}^k) \right\},$$

(4.32)

whereas, the death acceptance probability is

$$A(\varphi_{nk}^k, \varphi_{nk}^k \setminus \{x\}) = \min \left\{ 1, 1/r(\varphi_{nk}^k \setminus \{x\}, x) \right\} = \min \left\{ 1, \frac{1}{|\mathcal{W}|} \lambda^*(x|\varphi_{nk}^k \setminus \{x\})^{-1} \right\}.$$

(4.33)
The general form of the algorithm for random \( n \) is as follows.

**Algorithm:** M-H sampler for Markov MSTPPs

**Step 1:** Start with an initial valid state \( \varphi_{n_0}^0 \). For all \( k \geq 0 \), assume that at time \( k \) the Markov chain is at \( \varphi_{n_k}^k = \{x_1, \ldots, x_{n_k}\} \).

**Step 2:** With probability \( b(\varphi_{n_k}^k)(= 1/2) \) propose a birth or a death move.

**Step 3:** If a birth move is chosen, sample a point \( x \) from \( B(\varphi_{n_k}^k, x)(= 1/|X|) \) and accept the proposed transition to \( \varphi_{n_k}^{k+1} = \varphi_{n_k}^k \cup \{x\} \), with probability \( A(\varphi_{n_k}^k, \varphi_{n_k}^k \cup \{x\}) \). Go to step 5.

**Step 4:** If a death move is chosen, delete a point \( x \in \varphi_{n_k}^k \) with probability \( D(\varphi_{n_k}^k, x)(= 1/n) \), and accept the proposed transition to \( \varphi_{n_k}^{k+1} = \varphi_{n_k}^k \setminus \{x\} \), with probability \( A(\varphi_{n_k}^k, \varphi_{n_k}^k \setminus \{x\}) \), provided that \( \varphi_{n_k}^k \setminus \{x\} \neq \emptyset \). Go to step 5. If \( \varphi_{n_k}^k \setminus \{x\} = \emptyset \) we do nothing and go to step 1.

**Step 5:** (Change transition) We can include a “change move” for better mixing, i.e., choose a point \( x_j \) and replace it with a point \( u \sim c(\varphi_{n_k}^k, u, x_j) = 1/n|W_s|I(t_j = t_u)I(z_j = z_u) \), so that once we choose the \( j^{th} \) point we do not sample its time stamp or mark value, only the location component. As a result the M-H ratio is given by

\[
r(\varphi_{n_k}^k, x_k, u) = \frac{f((\varphi_{n_k}^k \setminus \{x_k\}) \cup \{u\})c(\varphi_{n_k}^k, u, x_k)}{f(\varphi_{n_k}^k)c(\varphi_{n_k}^k, x_k, u)} = \frac{\lambda^*(u|\varphi_{n_k}^k \setminus \{x_k\})}{\lambda^*(x_k|\varphi_{n_k}^k \setminus \{x_k\})}, \tag{4.34}
\]

and therefore

\[
r(\varphi_{n_k}^k, x_k, u) = \gamma^{R_n(\varphi_{n_k}^k, \rho)} e^H, \tag{4.35}
\]

where

\[
H = -\frac{1}{2}[(s_u - W(s_u)\beta - \phi_{t_u})^T \Sigma^{-1}(t_u)(s_u - W(s_u)\beta - \phi_{t_u})
- (s_k - W(s_k)\beta - \phi_{t_k})^T \Sigma^{-1}(t_k)(s_k - W(s_k)\beta - \phi_{t_k})], \tag{4.36}
\]
and

\[
R_n(\varphi_n^k, \rho) = \sum_{1 \leq i \leq n} I(\|s_u - s_i\| \leq \rho)I(|t_u - t_i| \leq 1)
- \sum_{1 \leq i \leq n \atop i \neq j} I(\|s_j - s_i\| \leq \rho)I(|t_j - t_i| \leq 1),
\]

(4.37)

counts the neighbors of the proposed point \(u\) minus the neighbours of the point to be removed \(x_j\). When we include change moves, we set the birth, death and change move probabilities to 1/3. Note that we could run the algorithm with change moves only, thus keeping \(n\) fixed. Consequently, \(\xi\) is fixed in this case and does not affect the number of points.

**Step 6:** If the proposed value is not accepted, remain at the previous state, i.e., \(\varphi_{n_k+1} = \varphi_{n_k}\).

Go to step 1.

### 4.5 Simulation study

Next we present some simulations in order to appreciate the multitude of point patterns that can be modeled under the setup of models 1-3. Recall that all models have common parameters \(\rho, \gamma, \xi, p\) and \(\beta\), and therefore, in all cases we use the same priors. In particular, we let \(\gamma|a_\gamma, b_\gamma \sim G(a_\gamma, b_\gamma), a_\gamma = 1, b_\gamma = 1\), and use an empirical Bayes prior for \(\xi|a_\xi, b_\xi \sim G(a_\xi, b_\xi)\), with \(a_\xi = 2, b_\xi = .5T|\mathcal{W}_s|/(n + 1)\). Given a square domain \(\mathcal{W}_s\) with side of size \(w\), we set \(\rho|\rho_{\text{max}} \sim Unif(0, \rho_{\text{max}})\), where \(\rho_{\text{max}} = \sqrt{2}w\). For random marks, we choose a Dirichlet prior for \(p|d \sim \text{Dirichlet}(d)\), with \(d = [1, ..., 1]^T\). Finally, for the coefficients \(\beta\) we set \(\beta|\mu_\beta, \sigma_\beta^2 \sim \mathcal{N}_{2p}(\mu_\beta, \sigma_\beta^2 I_{2p})\), where \(\mu_\beta = 0_{2p \times 1}\), and \(\sigma_\beta^2 = 0.1\), with \(p = 4\).

#### 4.5.1 Simulations for model 1

The parameters of model 1 include \(\theta = (\rho, \gamma, \xi, p, \beta, \theta_1)\), with \(\theta_1 = (\sigma_2^2, a, b)\). Table 4.1 presents the true values and posterior estimates for a (near) hard-core point pattern, across 10 time periods, with a random number of events at each time period. We run 20000
iterations of the MCMC (with a 10% burn-in period). Note that all of the true values are recovered well. Figures 4.1 and 4.2 present the events per time period, along with circles indicating the hard core distance ($\rho = 0.9$). Note that there is no dynamic evolution for any of the parameters of this model.

### 4.5.2 Simulations for model 2

The parameters of model 2 are $\theta = (\rho, \gamma, \xi, \mathbf{p}, \beta, \theta_1)$, with $\theta_1 = (\Sigma, \{\phi_t\}_{t=1}^n)$. The true values of the model and process parameters are obtained by sampling from their corresponding prior distributions or using their evolution model in the case of $\{\phi_t\}_{t=1}^n$. In particular, we choose $\Sigma|\nu_\Sigma, C_\Sigma \sim \text{Wishart}_2(\nu_\Sigma, (\nu_\Sigma C_\Sigma)^{-1})$, with $\nu_\Sigma = 4$, and $C_\Sigma = \text{diag}(0.1, 0.1)$. For the dynamic evolution model for $\phi_t|\phi_{t-1}, \Phi, \Sigma_\phi$, we set $\Sigma_\phi|\nu_\phi, C_\phi \sim \text{Wishart}_2(\nu_\phi, (\nu_\phi C_\phi)^{-1})$, with $\nu_\phi = 4$, and $C_\phi = \text{diag}(0.1, 0.1)$, and model the auto-covariance function $a \text{ priori}$ using $\text{vec}(\mathbf{\Phi})| \mu_\Phi, \Sigma_\Phi \sim \mathcal{N}_4(\mu_\Phi, \Sigma_\Phi)$, where $\mu_\Phi = 0_{4 \times 1}$, and $\Sigma_\Phi = \text{diag}(0.1, 0.1, 0.1, 0.1)$.

We present two runs of model 2 for different values of the parameter $\gamma$ in Figures 4.3 and 4.4, respectively. Note that the second model has a fixed number of events at each time period and the events are never closer than $\rho = 1$ distance (hard-core model). We note that all parameters are recovered using the Bayesian MCMC sampler, as illustrated in Tables 4.2 and 4.3.

### 4.5.3 Simulations for model 3

The parameters of model 3 include $\theta = (\rho, \gamma, \xi, \mathbf{p}, \beta, \theta_1)$, with $\theta_1 = (\{\phi_t\}_{t=1}^n, \{\Sigma(t)\}_{t=1}^n)$, and the true values are obtained by sampling from their prior distributions or using their evolution models. The priors and corresponding hyper-parameters are the same as those in model 2, but now instead of $\Sigma$ we model $\Sigma(t)$ using process hyper-parameters $\tilde{\mu}_{\kappa_1} = 0$, $\tilde{\sigma}_{\kappa_1}^2 = 0.1$, $\tilde{\mu}_{\kappa_2} = 0$, $\tilde{\sigma}_{\kappa_2}^2 = 0.1$, $\tilde{\mu}_\zeta = 0$, $\tilde{\sigma}_\zeta^2 = 0.1$. Finally, we take fixed hyper-parameters $\sigma_\zeta^2 = 0.1$, $\sigma_{\kappa_1}^2 = 0.1$, $\sigma_{\kappa_2}^2 = 0.1$, and assume a random walk for the $\zeta$, $\kappa_1$, and $\kappa_2$, by setting $\varphi_\zeta = \varphi_{\kappa_1} = \varphi_{\kappa_2} = 1$. 

175
We present a run of model 3 with 5 events at each time period, for 5 time periods, and for a near hard-core model with $\gamma = 0.01$. The patterns are displayed in Figure 4.5, where we can clearly see that, in most cases, the events are never closer than $\rho = 4$ distance (hard-core model). Finally, we present some of the parameter estimates in Table 4.4, and notice that the posterior means are close to the true values.

### 4.6 Application to modeling tornado incidence

The National Oceanic and Atmospheric Administration (NOAA, [http://www.noaa.gov/](http://www.noaa.gov/)) is a U.S. agency tasked with the dissemination of daily weather forecasts and severe storm warnings. The Storm Prediction Center of NOAA archives important information on tornado occurrences throughout the U.S., starting from 1950 up to the present. We will consider events for the years 2005-2010 and assess our point process models for the observed events.

The data consists of the starting location (where the tornado touched the ground) in longitude-latitude as well as its destructive power in the “Fujita scale” F0-F5, which will be treated as the marks of the events. The Fujita scale is a rating system for tornado damage in which F0 represents “minimal damage” and F5 represents “complete destruction.” In particular, F0 and F1 consist of recorded tornadoes that have barely touched the ground, whereas, F2-F5 tornadoes cause significant damage on the ground.

There are other covariates/marks of interest in the data, however, we only consider the F values as marks associated with each event. In particular, the mark is a polytomous variable with 6 discrete marks (valued 0-5), and we will consider only the more intense tornadoes with range F2-F5. For more details on the Fujita scale and its properties we refer to [Wikle and Anderson, 2003](http://www.noaa.gov/), and the references therein). Note that since 2007, the USA has adopted the EF (Enhanced Fujita) scale, which is an enhancement of the original F-scale that results in a more robust and standardized measure of tornado intensity.

Figure 4.6 presents the tornado events from years 2005 to 2010 (623 events in total).
that each event location is displayed as a disc of different sizes depending on the F-level of the event.

### 4.6.1 Results from model 1

Table 4.5 presents the posterior means and 95% credible sets of all the model parameters. The interaction distance is estimated at $\hat{\rho} = 23.8104$, with 95% credible set given by $[20.37702, 27.70137]$. Note that for this model $\hat{\gamma} = 0.9664598$ is close to 1, with 95% credible set $[0.9652222, 0.9674125]$, indicating that although the point patterns are near independence (the case $\gamma = 1$), there is weak interaction between the events of the point patterns, between and within time period.

Moreover, the posterior means of the mark distribution probabilities are given by $\hat{p}_1 = 0.7048503$ (95% CS $[0.6687929, 0.7405099]$), $\hat{p}_2 = 0.2440887$ (95% CS $[0.2114171, 0.2781614]$), $\hat{p}_3 = 0.04629898$ (95% CS $[0.03111901, 0.064273]$), and $\hat{p}_4 = 0.004761964$ (95% CS $[0.0010181, 0.01158546]$) for observing tornadoes with powers F2-F5, respectively.

### 4.6.2 Results from model 2

Table 4.6 presents the posterior means and 95% credible sets of all the model parameters. In contrast with model 1, model 2 yields an estimator of the interaction distance given by $\hat{\rho} = 0.1374936$, with 95% credible set given by $[0.1143008, 0.1713605]$, and a point estimator $\hat{\gamma} = 4.370644$, with 95% credible set $[3.261369, 5.889819]$. Therefore, since in model 2 the means of the main effects are allowed to evolve over time, the fit indicates clustering of the events across time and space (which was not identified by model 1).

Certain parameters of the model are not affected by the specific evolution model chosen for the main effect means, e.g., the number of events controlled by $\xi$ is estimated by similar values from both models, and the mark distribution probabilities are estimated with almost identical values for both models 1 and 2.
4.6.3 Results from model 3

Table 4.7 presents the posterior means and 95% credible sets of all the model parameters. In contrast with models 1 and 2, model 3 yields an estimator of the interaction distance given by $\hat{\rho} = 2.206598$, with 95% credible set given by $[0.2071125, 4.490314]$, and a point estimator $\hat{\gamma} = 1.087929$, with 95% credible set $[0.0444709, 2.855626]$. Therefore, since in model 3 the means of the main effects, as well as, the covariance matrices are allowed to evolve over time, the fit indicates once again clustering of the events across time and space, but not as strong as model 2.

Once again we notice that certain parameters of the model are not affected by the specific evolution model chosen for the main effect means and covariances, e.g., the number of events controlled by $\xi$ is estimated by similar values in all models fit, and the mark distribution probabilities are estimated with almost identical values for both models 1, 2 and 3.

4.7 Discussion

We presented a hierarchical Bayesian framework for modeling the joint distribution of a marked Markov space-time point process model, while taking care of potential interactions between the events. The three models presented can be used in different situations depending on the type of data under consideration.

In particular, model 1 allows the main effect means to evolve over time, but none of the other parameters are evolving dynamically. This model would be more appropriate in modeling fast evolving locations of the events, for example, a herd of animals as it moves fast across the domain of observation, with the animals staying close to each other (i.e., fairly clustered).

Model 2 allows the main effects to evolve over time, while keeping the covariance matrix (the layout of the locations) fixed across time, which can be useful in modeling all the point patterns under model 1, but in addition, allow us to capture a more complicated variability structure.
Moreover, model 3 can be thought of as an extension of models 1 and 2, capturing both cases, but now model 3 allows even the covariances to evolve over time. These models can be particularly useful when modeling events with interactions and layout structure that varies over time.

In all three models, the value of the parameter $\gamma$ can be used to model different scenarios. For example, a hard-core model for $\gamma = 0$, an inhibition model for $0 < \gamma < 1$, a Poisson pattern for $\gamma = 1$ (no interaction effects) and finally a model for clustering for $\gamma > 1$. Unlike classic models like the Strauss model, the proposed model has a valid joint distribution in the case $\gamma > 1$, which is an improvement in models with Strauss like-interaction functions used to model point patterns exhibiting clustering.

Finally, in terms of the tornado data, all three models appear to be arriving at similar estimators for some of their common parameters, however, they differ for the important parameters $\gamma$ and $\rho$. If we were to choose one of the three models as the more appropriate, we would choose model 3, since its additional complexity allows it to capture the process dynamic better, and in addition, model 3 contains the other two models as special cases.
4.A Model 1: Posterior distribution and full conditionals

The full posterior distribution $\theta | \varphi_n$ of model 1, where $\theta = (\rho, \gamma, \xi, p, \beta, \theta_1)$, with $\theta_1 = (\sigma^2, a, b)$ and $\phi_n = t_a + b$, is given by

$$
\pi(\theta, \eta | \varphi_n) \propto f_n(\varphi_n | \theta) \pi(\theta | \eta) \quad (4.38)
$$

$$
\propto \prod_{i=1}^{n} \frac{p_{zi}}{c_n(\theta)} \xi^{n_{i} \gamma} \sum_{1 \leq i \neq j \leq n} I(||s_i - s_j|| \leq \rho) I(||t_i - t_j|| \leq 1) \!
\times \! I_{[0, \rho_{\max}]}(\rho) \xi^{a\gamma - 1} e^{-\xi/b\xi} \gamma^{a\gamma - 1} e^{-\gamma/b\gamma} \frac{1}{(\sigma^2)^{a\alpha + 1}} e^{-\frac{1}{2\sigma^2} \sum_{j=1}^{M} p_j^{d_j - 1}} \prod_{j=1}^{M} p_j^{d_j - 1} \!
\times e^{-\frac{1}{2\sigma^2} ||a - \mu_a||^2} e^{-\frac{1}{2\sigma^2} ||b - \mu_b||^2} e^{-\frac{1}{2\sigma^2} ||\beta - \mu_\beta||^2},
$$

with full conditionals

$$
\pi(p_j | .) \propto \prod_{j=1}^{M} p_j^{k_j + d_j - 1}, \quad k_j = \#(z_i = j),
$$

$$
\pi(\rho | .) \propto 1 \frac{1}{c_n(\theta)} \gamma^{1\sum_{1 \leq i \neq j \leq n} I(||s_i - s_j|| \leq \rho) I(||t_i - t_j|| \leq 1) \!
\times \! I_{[0, \rho_{\max}]}(\rho),
$$

$$
\pi(\gamma | .) \propto 1 \frac{1}{c_n(\theta)} \gamma^{a\gamma - 1 + \sum_{1 \leq i \neq j \leq n} I(||s_i - s_j|| \leq \rho) I(||t_i - t_j|| \leq 1) \!
\times \! e^{-\gamma/b\gamma},
$$

$$
\pi(\sigma^2 | .) \propto 1 \frac{1}{c_n(\theta)} (\sigma^2)^{-a\sigma - 1} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} ||s_i - W(s_i)\beta - \phi_i||^2},
$$

$$
\pi(\xi | .) \propto 1 \frac{1}{c_n(\theta)} \xi^{a\xi + n - 1} e^{-\xi/b\xi},
$$

with $a_\xi = 1$, $b_\xi = \frac{n}{(a_\xi + n)|W_n|}$, and

$$
\pi(a | .) \propto 1 \frac{1}{c_n(\theta)} e^{-\frac{1}{2}(a - \mu_a)^T \Sigma_a^{-1}(a - \mu_a)},
$$

$$
\pi(b | .) \propto 1 \frac{1}{c_n(\theta)} e^{-\frac{1}{2}(b - \mu_b)^T \Sigma_b^{-1}(b - \mu_b)},
$$

and

$$
\pi(\beta | .) \propto 1 \frac{1}{c_n(\theta)} e^{-\frac{1}{2}(\beta - \mu_\beta)^T \Sigma_\omega^{-1}(\beta - \mu_\omega)}.
$$
with
\[
\mu_a = \Sigma_a \left( \frac{1}{\sigma^2} \sum_{i=1}^{n} t_i [s_i - W(s_i)^T \beta] - \frac{1}{\sigma^2} \sum_{i=1}^{n} t_i b + \frac{1}{\sigma^2} \mu_a \right),
\] (4.47)
\[
\Sigma_a = \left[ \frac{1}{\sigma^2} \sum_{i=1}^{n} t_i^2 + 1/\sigma^2 \right]^{-1} I_2,
\] (4.48)
\[
\mu_b = \Sigma_b \left( \frac{1}{\sigma^2} \sum_{i=1}^{n} [s_i - W(s_i)^T \beta] - \frac{1}{\sigma^2} \sum_{i=1}^{n} t_i a + \frac{1}{\sigma^2} \mu_b \right),
\] (4.49)
\[
\Sigma_b = \left[ \frac{n}{\sigma^2} + 1/\sigma^2 \right]^{-1} I_2,
\] (4.50)
\[
\mu_W = \Sigma_W \left( \frac{1}{\sigma^2} \sum_{i=1}^{n} W(s_i)^T [s_i - \phi_t] + \frac{1}{\sigma^2} \mu_\beta \right),
\] (4.51)
\[
\Sigma_W = \left[ \frac{1}{\sigma^2} \sum_{i=1}^{n} W(s_i)^T W(s_i) + \frac{1}{\sigma^2} I_{2p} \right]^{-1},
\] (4.52)

where \( \Sigma_W \) is a \( 2p \times 2p \) matrix and \( W(s) \) is a \( 2 \times 2p \) matrix. Note that in our examples we take \( p = 4 \). Furthermore, the proposals for the parameters involving the intractable constant term \( \frac{1}{c_n(\theta)} \) in double M-H sampler, are generated from the distributions without the term \( \frac{1}{c_n(\theta)} \). For example, the proposed value for \( \sigma^2 \) is generated according to an inverse gamma with shape parameter \( a_\sigma \) and scale parameter \( \left[ \frac{1}{b_\sigma} + \frac{1}{2} \sum_{i=1}^{n} ||s_i - W(s_i)\beta - \phi_t||^2 \right]^{-1} \) (that is, we ignore the \( \frac{1}{c_n(\theta)} \) term in (4.42)).

### 4.B Model 2: Posterior distribution and full conditionals

The full posterior distribution \( \theta | \varphi_n \) of model 2, where \( \theta = (\rho, \gamma, \xi, p, \beta, \theta_1) \), with \( \theta_1 = (\Sigma, \{\phi_t\}_{i=1}^{n}) \), is given by
\[
\pi(\theta, \eta | \varphi_n) \propto f_n(\varphi_n | \theta) \pi(\theta | \eta)
\] (4.53)
\[
\propto \prod_{i=1}^{n} \frac{p_{z_i}}{c_n(\theta)} \xi^{n_{\gamma_j} \leq n} I_{\{\|s_i - s_j\| \leq \rho\}} I_{\{||t_i - t_j|| \leq 1\}} |\Sigma|^{-n/2}
\times e^{-\frac{1}{2} \sum_{i=1}^{n} (s_i - W(s_i)\beta - \phi_t)^T \Sigma^{-1} (s_i - W(s_i)\beta - \phi_t)}
\times \left( \prod_{j=1}^{M} p_{j-1} \right) I_{[0, \rho_{\text{max}}]}(\rho) \xi^{a_{\gamma_j} + 1 - \xi/b_{\gamma_j} a_{\gamma_j} - 1 - \epsilon/\gamma_j b_{\gamma_j} |\Sigma|^{-(\nu_\gamma + 1)/2} e^{-\frac{\gamma_j}{2} \text{tr}(C_{\gamma} \Sigma^{-1})}
\times e^{-\frac{1}{2\beta} ||\beta-\mu_\beta||^2} e^{-\frac{1}{2} (\phi_0 - \mu_0)^T \Sigma_0^{-1} (\phi_0 - \mu_0)} |\Sigma_0|^{-T/2} e^{-\frac{1}{2} \sum_{i=1}^{T} (\phi_t - \Phi_{t-1})^T \Sigma_{\phi_t}^{-1} (\phi_t - \Phi_{t-1})}
\times |\Sigma_\phi|^{-1/2} e^{-\frac{1}{2} (\text{vec}(\Phi) - \mu_\Phi)^T \Sigma_\phi^{-1} (\text{vec}(\Phi) - \mu_\Phi)} |\Sigma_\phi|^{-(\nu_\phi + 1)/2} e^{-\frac{\nu_\phi}{2} \text{tr}(C_{\phi} \Sigma_\phi^{-1})}
\times 181
\]
and the full conditionals for $\mathbf{p}$, $\rho$, $\gamma$, and $\xi$ are the same as in the previous model given by equations (4.39), (4.40), (4.41) and (4.43). In addition,

$$
\pi(\mathbf{p}|\cdot) \propto \frac{1}{c_n(\theta)} |\Sigma|^{-(n+\nu_\Sigma+2+1)/2} e^{-\frac{1}{2} tr \left( \nu_\Sigma C_{\nu} + \sum_{i=1}^{n}(s_i - W(s_i)\beta - \phi_i)(s_i - W(s_i)\beta - \phi_i)^T \right) \Sigma^{-1}},
$$

(4.54)

$$
\pi(\rho|\cdot) \propto \frac{1}{c_n(\theta)} e^{-\frac{1}{2} (\beta - \mu_W)^T \Sigma_W^{-1} (\beta - \mu_W)},
$$

(4.55)

with

$$
\mu_W = \Sigma_W \left( \sum_{i=1}^{n} W(s_i)^T \Sigma^{-1} (s_i - \phi_i) + \frac{1}{\sigma_\beta} \mu_0 \right),
$$

and

$$
\Sigma_W = \left[ \sum_{i=1}^{n} W(s_i)^T \Sigma^{-1} W(s_i) + \frac{1}{\sigma_\beta^2} I_{2p} \right]^{-1},
$$

where $\mu_W$ is a $2p \times 1$ vector and $\Sigma_W$ is a $2p \times 2p$ matrix.

Now let

$$
n_t = \#(i : t_i = t),
$$

(4.56)

denote the number of events at time $t = 1, \ldots, T$. Then, it is straightforward to see that

$$
\pi(\phi_0|\cdot) \propto \frac{1}{c_n(\theta)} |A_0|^{-\frac{1}{2}} e^{-\frac{1}{2} (\phi_0 - A_0a_0)^T A_0^{-1} (\phi_0 - A_0a_0)},
$$

(4.57)

with

$$
a_0 = \Phi^T \Sigma^{-1}_\phi \phi_1 + \Sigma^{-1}_0 \mu_0,
$$

and

$$
A_0 = (\Phi^T \Sigma^{-1}_\phi \Phi + \Sigma^{-1}_0)^{-1}.
$$

For $t = 1, \ldots, T - 1$, we have

$$
\pi(\phi_t|\cdot) \propto \frac{1}{c_n(\theta)} |A_t|^{-\frac{1}{2}} e^{-\frac{1}{2} (\phi_t - A_ta_t)^T A_t^{-1} (\phi_t - A_ta_t)},
$$

(4.58)

with

$$
a_t = \sum_{i : t_i = t} \Sigma^{-1}_\phi (s_i - W(s_i)\beta) + \Phi^T \Sigma^{-1}_\phi \phi_{t+1} + \Sigma^{-1}_\phi \Phi \phi_{t-1},
$$

and for $t = T$, we obtain

$$
\pi(\phi_T|\cdot) \propto \frac{1}{c_n(\theta)} |A_T|^{-\frac{1}{2}} e^{-\frac{1}{2} (\phi_T - A_Ta_T)^T A_T^{-1} (\phi_T - A_Ta_T)},
$$

(4.59)

with

$$
A_T = (\Sigma^{-1}_\phi + n_T \Sigma^{-1})^{-1},
$$
and
\[ a_T = \sum_{i:t_i=T} \Sigma^{-1} (s_i - W(s_i)\beta) + \Sigma^{-1}_{\phi} \Phi_{T-1}. \]

Moreover,
\[
\pi(\Sigma_{\phi}, \cdot) \propto |\Sigma_{\phi}|^{-((\nu_0+T)+2)/2} e^{-\frac{1}{2} \text{tr}\left( \left( \sum_{t=1}^{T} (\phi_t - \Phi)\Phi_{T-1})^{T} + \nu_{\phi} C_{\phi} \right) \Sigma_{\phi}^{-1} \right)} \tag{4.60}
\]
and letting \( \Phi = \text{vec}(\Phi) \), we have
\[
\pi(\Phi, \cdot) \propto |A_{\Phi}|^{-\frac{1}{2}} e^{-\frac{1}{2} (\Phi - A_{\Phi} a_{\Phi})^{T} A_{\Phi}^{-1} (\Phi - A_{\Phi} a_{\Phi})} \tag{4.61}
\]
with
\[
A_{\Phi} = (\mathcal{Y}_{0:T-1} \otimes I_2)^T (I_T \otimes \Sigma_{\phi})^{-1} (\mathcal{Y}_{0:T-1} \otimes I_2) + \Sigma_{\phi}^{-1},
\]
\[
a_{\Phi} = (\mathcal{Y}_{1:T} \otimes I_2)^T (I_T \otimes \Sigma_{\phi})^{-1} \text{vec}(\mathcal{Y}_{1:T}) + \Sigma_{\phi}^{-1} \mu_{\Phi},
\]
where \( \mathcal{Y}_{0:T-1} = [\phi_0, \ldots, \phi_{T-1}], \mathcal{Y}_{1:T} = [\phi_1, \ldots, \phi_{T}] \) are \( 2 \times T \) matrices. The hyper-parameters \( \mu_0, \Sigma_0, \nu_{\Sigma}, \nu_{\phi}, \mathcal{C}_{\phi}, \mu_{\Phi} \) and \( \Sigma_{\Phi} \) are assumed to be fixed.

4.C Model 3: Posterior distribution and full conditionals

The full posterior distribution \( \theta|\varphi_n \) of model 3, where \( \theta = (\rho, \gamma, \xi, p, \beta, \theta_1) \), with \( \theta_1 = (\{\phi_t\}_{t=1}^{n}, \{\Sigma(t_i)\}_{t=1}^{n}) \), is given by
\[
\pi(\theta, \eta|\varphi_n) \propto f_n(\varphi_n|\theta) \pi(\theta|\eta) \tag{4.62}
\]
\[
\propto \prod_{i=1}^{n} p_{z_i} \prod_{i=1}^{n} \frac{\gamma_{i|\theta}}{c_1(\theta)^{\frac{1}{2}}} \sum_{\sum_{i=1}^{n} I(||s_i - s_j|| \leq \rho) I(t_i - t_j) \leq 1} \left[ \prod_{i=1}^{n} \left| \Sigma(t_i) \right|^{-\frac{1}{2}} \right]
\times e^{-\frac{1}{2} \sum_{i=1}^{n} (s_i - W(s_i)\beta - \phi_t)^T \Sigma(t_i)^{-1} (s_i - W(s_i)\beta - \phi_t) - \frac{1}{2} \sum_{i=1}^{n} (\phi_t - \Phi(t_i))^{T} \Sigma_{\phi}^{-1} (\phi_t - \Phi(t_i))}
\times e^{-\frac{1}{2} (\rho - \mu_0)^T \Sigma_{\phi}^{-1} (\rho - \mu_0) / 2} e^{-\frac{1}{2} \sum_{t=1}^{T} (\phi_t - \Phi(t_i))^{T} \Sigma_{\phi}^{-1} (\phi_t - \Phi(t_i))}
\times \left| \Sigma_{\Phi} \right|^{-1/2} e^{-\frac{1}{2} (\text{vec}(\Phi) - \mu_{\Phi})^{T} \Sigma_{\phi}^{-1} (\text{vec}(\Phi) - \mu_{\Phi})} \left| \Sigma_{\phi} \right|^{-(\nu_{\phi}+2)/2} e^{-\nu_{\phi} \text{tr}(\mathcal{C}_{\phi} \Sigma_{\phi}^{-1})}
\times \prod_{i=1}^{n} \frac{1}{\sigma_{\kappa_1}} e^{-\frac{1}{2\sigma_{\kappa_1}^2} (\log(\kappa_1(t)) - \tilde{\kappa}_1)^2} T \prod_{t=1}^{T} \frac{1}{\sigma_{\kappa_1}} e^{-1} \prod_{t=1}^{T} \frac{1}{\sigma_{\kappa_2}} e^{-\frac{1}{2\sigma_{\kappa_2}^2} (\log(\kappa_2(t)) - \tilde{\kappa}_2)^2} \prod_{t=1}^{T} \frac{1}{\sigma_{\kappa_2}} e^{-1}
\]
\[
183
\]
\[
\frac{1}{\sigma_{\kappa_1}} \frac{1}{\sigma_{\kappa_1} + 1} \prod_{t=1}^{T} \frac{1}{\sigma_{\kappa_t}} \frac{1}{\sigma_{\kappa_t} + 1} e^{-\frac{1}{2}(\zeta(t) - \varphi \zeta(t-1))^2} \times \frac{1}{\sigma_{\kappa_t}} \frac{1}{\sigma_{\kappa_t} + 1} \prod_{t=1}^{T} \frac{1}{\sigma_{\kappa_t}} \frac{1}{\sigma_{\kappa_t} + 1} e^{-\frac{1}{2}(\zeta(t) - \varphi \zeta(t-1))^2} \\
\times \int_{\kappa_1, d_{\kappa_1}} (\varphi_{\kappa_1})^I_{\kappa_1, d_{\kappa_2}} (\varphi_{\kappa_2})^I_{\kappa_2, d_{\kappa_3}} (\varphi_{\kappa_3}) \\
\text{and the full conditionals for } \mathbf{p}, \rho, \gamma, \text{ and } \xi \text{ are the same as in model 1 given by equations (4.39), (4.40), (4.41) and (4.43). In addition,} \\
\pi(\beta|.) \propto \frac{1}{c_n(\theta)} e^{-\frac{1}{2}(\beta - \mu W)^T \Sigma W^{-1}(\beta - \mu W)} , \tag{4.63}
\]

with \[
\mu_W = \Sigma W \left( \sum_{i=1}^{n} W(s_i)^T \Sigma(t_i)^{-1} (s_i - \phi_{t_i}) + \frac{1}{\sigma_{\beta}} \mu_{\beta} \right) ,
\]
and \[
\Sigma_W = \left[ \sum_{i=1}^{n} W(s_i)^T \Sigma(t_i)^{-1} W(s_i) + \frac{1}{\sigma_{\beta}} I_{2p} \right]^{-1}.
\]

where \( \mu_W \) is a \( 2p \times 1 \) matrix and \( \Sigma_W \) is a \( 2p \times 2p \) matrix.

Now with \( n_t \) as in (4.56), we have \[
\pi(\phi_0|. ) \propto \frac{1}{c_n(\theta)} |A_0|^{-\frac{1}{2}} e^{-\frac{1}{2}(\phi_0 - A_0 a_0)^T A_0^{-1}(\phi_0 - A_0 a_0)} , \tag{4.64}
\]
with \[
A_0 = (\Phi^T \Sigma^{-1}_{\phi} + \Sigma^{-1}_{\phi})^{-1}.
\]

For \( t = 1, \ldots, T - 1 \), we have \[
\pi(\phi_t|. ) \propto \frac{1}{c_n(\theta)} |A_t|^{-\frac{1}{2}} e^{-\frac{1}{2}(\phi_t - A_t a_t)^T A_t^{-1}(\phi_t - A_t a_t)} , \tag{4.65}
\]
with \[
A_t = (\Phi^T \Sigma^{-1}_{\phi} + \Sigma^{-1}_{\phi} + n_t \Sigma(t)^{-1})^{-1} , \\
a_t = \sum_{i:t_i=t} \Sigma(t_i)^{-1} (s_i - W(s_i) \beta) + \Phi^T \Sigma^{-1}_{\phi} \phi_{t+1} + \Sigma^{-1}_{\phi} \phi_{t-1} ,
\]
and for \( t = T \), we obtain \[
\pi(\phi_T|. ) \propto \frac{1}{c_n(\theta)} |A_T|^{-\frac{1}{2}} e^{-\frac{1}{2}(\phi_T - A_T a_T)^T A_T^{-1}(\phi_T - A_T a_T)} , \tag{4.66}
\]
with \[
A_T = (\Sigma^{-1}_{\phi} + n_T \Sigma(T)^{-1})^{-1} .
\]
and
\[ a_T = \sum_{i:t_i = T} \Sigma(t_i)^{-1} (s_i - W(s_i)\beta) + \Sigma^{-1}_\phi \phi_{T-1}. \]

The full-conditionals for \( \Sigma_\phi \) and \( \tilde{\Phi} = \text{vec}(\Phi) \) are given by (4.60) and (4.61), respectively. Recall that \( \Sigma(t) = K(t)\Gamma(t)\Gamma(t)^T K(t) \), with \( K(t) = \text{diag}(\kappa_1(t), \kappa_2(t)) \) and \( \Gamma(t) = \begin{pmatrix} 1 & 0 \\ \zeta(t) & 1 \end{pmatrix} \). The full conditional distributions of the elements of \( \Sigma(t) \) are given next.

**The full conditional distribution of \( \kappa_1(t) \)**
For \( t = 0 \), we have
\[ \pi(\kappa_1(0)|.) \propto \exp \left\{ -\frac{\sigma^2_{\kappa_1} + \varphi^2_{\kappa_1} \sigma^2_{\kappa_1}}{2\sigma^2_{\kappa_1} \sigma^2_{\kappa_1}} \left( \log(\kappa_1(0)) - \frac{-\bar{\mu}_{\kappa_1} \sigma^2_{\kappa_1} + \varphi_{\kappa_1} \sigma^2_{\kappa_1} \log(\kappa_1(1))}{\sigma^2_{\kappa_1} + \varphi_{\kappa_1} \sigma^2_{\kappa_1}} \right)^2 \right\}, \quad (4.67) \]
for \( t = 1, \ldots, T - 1, \)
\[ \pi(\kappa_1(t)|.) \propto |\Sigma(t)|^{-\frac{n_t}{2}} \times \exp \left\{ -\frac{1}{2} \sum_{i:t_i = t} (s_i - W(s_i)\beta - \phi_{t_i})^T \Sigma^{-1}(t_i) (s_i - W(s_i)\beta - \phi_{t_i}) \right\} \quad (4.68) \]
\[ \times \exp \left\{ -\frac{1 + \varphi^2_{\kappa_1}}{2\sigma^2_{\kappa_1}} \left( \log(\kappa_1(t)) - \frac{\varphi_{\kappa_1}(\log(\kappa_1(t-1)) + \log(\kappa_1(t+1)))}{1 + \varphi^2_{\kappa_1}} \right)^2 \right\}, \]
and for \( t = T, \)
\[ \pi(\kappa_1(T)|.) \propto |\Sigma(T)|^{-\frac{n_T}{2}} \times \exp \left\{ -\frac{1}{2} \sum_{i:t_i = T} (s_i - W(s_i)\beta - \phi_{t_i})^T \Sigma^{-1}(t_i) (s_i - W(s_i)\beta - \phi_{t_i}) \right\} \quad (4.69) \]
\[ \times \exp \left\{ -\frac{1}{2\sigma^2_{\kappa_1}} (\log(\kappa_1(T)) - \varphi_{\kappa_1} \log(\kappa_1(T-1))^2 \right\}. \]

**The full conditional distribution of \( \kappa_2(t) \)**
For \( t = 0 \), we have
\[ \pi(\kappa_2(0)|.) \propto \exp \left\{ -\frac{\sigma^2_{\kappa_2} + \varphi^2_{\kappa_2} \sigma^2_{\kappa_2}}{2\sigma^2_{\kappa_2} \sigma^2_{\kappa_2}} \left( \log(\kappa_2(0)) - \frac{-\bar{\mu}_{\kappa_2} \sigma^2_{\kappa_2} + \varphi_{\kappa_2} \sigma^2_{\kappa_2} \log(\kappa_2(1))}{\sigma^2_{\kappa_2} + \varphi_{\kappa_2} \sigma^2_{\kappa_2}} \right)^2 \right\}, \quad (4.70) \]
for \( t = 1, \ldots, T - 1, \)
\[ \pi(\kappa_2(t)|.) \propto |\Sigma(t)|^{-\frac{n_t}{2}} \times \exp \left\{ -\frac{1}{2} \sum_{i:t_i = t} (s_i - W(s_i)\beta - \phi_{t_i})^T \Sigma^{-1}(t_i) (s_i - W(s_i)\beta - \phi_{t_i}) \right\} \quad (4.71) \]
\[ \times \exp \left\{ -\frac{1 + \varphi^2_{\kappa_2}}{2\sigma^2_{\kappa_2}} \left( \log(\kappa_2(t)) - \frac{\varphi_{\kappa_2}(\log(\kappa_2(t-1)) + \log(\kappa_2(t+1)))}{1 + \varphi^2_{\kappa_2}} \right)^2 \right\}, \]
and for $t = T$,

$$\pi(\kappa_2(T)|.) \propto |\Sigma(T)|^{-\frac{n_T}{2}} \times \exp \left\{ -\frac{1}{2} \sum_{i:t_i=T} (s_i - W(s_i)\beta - \phi_{t_i})^T \Sigma^{-1}(t_i)(s_i - W(s_i)\beta - \phi_{t_i}) \right\} \times \exp \left\{ -\frac{1}{2\sigma^2_{\kappa_2}} (\log(\kappa_2(T)) - \varphi_{\kappa_2} \log(\kappa_2(T) - 1))^2 \right\}. \tag{4.72}$$

The full conditional distribution of $\zeta(t_i)$

For $t = 0$, we have

$$\pi(\zeta(0)|.) \propto \exp \left\{ -\frac{\sigma^2_{\zeta} + \varphi^2_{\zeta}s^2_{\zeta}}{2\sigma^2_{\zeta}s^2_{\zeta}} \left( \zeta(0) - \frac{\mu_{\zeta}\sigma^2_{\zeta} + \varphi_{\zeta}\tilde{\sigma}^2_{\zeta}(1)}{\sigma^2_{\zeta} + \varphi_{\zeta}\tilde{\sigma}^2_{\zeta}} \right)^2 \right\}, \tag{4.73}$$

for $t = 2, \ldots, T - 1$,

$$\pi(\zeta(t)|.) \propto |\Sigma(t)|^{-\frac{n_t}{2}} \times \exp \left\{ -\frac{1}{2} \sum_{i:t_i=t} (s_i - W(s_i)\beta - \phi_{t_i})^T \Sigma^{-1}(t_i)(s_i - W(s_i)\beta - \phi_{t_i}) \right\} \times \exp \left\{ -\frac{1 + \varphi^2_{\zeta}}{2\sigma^2_{\zeta}} \left( \zeta(t) - \frac{\varphi_{\zeta}(\zeta(t) + (t + 1) + \zeta(t + 1))}{1 + \varphi^2_{\zeta}} \right)^2 \right\}, \tag{4.74}$$

and for $t = T$,

$$\pi(\zeta(T)|.) \propto |\Sigma(T)|^{-\frac{n_T}{2}} \times \exp \left\{ -\frac{1}{2} \sum_{i:t_i=T} (s_i - W(s_i)\beta - \phi_{t_i})^T \Sigma^{-1}(t_i)(s_i - W(s_i)\beta - \phi_{t_i}) \right\} \times \exp \left\{ -\frac{1}{2\sigma^2_{\zeta}} (\zeta(T) - \varphi_{\zeta}\zeta(T - 1))^2 \right\}. \tag{4.75}$$

Full conditional distributions for the hyper-parameters

It is straightforward to show that

$$\sigma^2_{\kappa_1}|. \sim IG \left( \frac{n + 2a_{\kappa_1}}{2}, \left[ \frac{1}{2} \sum_{t=2}^{T} \log(\kappa_1(t)) - \varphi_{\kappa_1} \log(\kappa_1(t - 1)))^2 + 1/b_{\kappa_1} \right]^{-1} \right), \tag{4.76}$$

$$\sigma^2_{\kappa_2}|. \sim IG \left( \frac{n + 2a_{\kappa_2}}{2}, \left[ \frac{1}{2} \sum_{t=2}^{T} \log(\kappa_2(t)) - \varphi_{\kappa_2} \log(\kappa_2(t - 1)))^2 + 1/b_{\kappa_2} \right]^{-1} \right), \tag{4.77}$$

$$\sigma^2_{\zeta}|. \sim IG \left( \frac{n + 2a_{\zeta}}{2}, \left[ \frac{1}{2} \sum_{t=2}^{T} (\zeta(t) - \varphi_{\zeta}\zeta(t - 1))^2 + 1/b_{\zeta} \right]^{-1} \right), \tag{4.78}$$
\[ \varphi_{\kappa_1} \sim \mathcal{N} \left( \frac{\sum_{t=2}^{T} \log(\kappa_1(t)) \log(\kappa_1(t-1))}{\sum_{t=2}^{T} (\log(\kappa_1(t-1)))^2}, \frac{\sigma^2_{\kappa_1}}{\sum_{t=2}^{T} (\log(\kappa_1(t-1)))^2} \right), \tag{4.79} \]

\[ \varphi_{\kappa_2} \sim \mathcal{N} \left( \frac{\sum_{t=2}^{T} \log(\kappa_2(t)) \log(\kappa_2(t-1))}{\sum_{t=2}^{T} (\log(\kappa_2(t-1)))^2}, \frac{\sigma^2_{\kappa_2}}{\sum_{t=2}^{T} (\log(\kappa_2(t-1)))^2} \right), \tag{4.80} \]

and

\[ \varphi_{\zeta} \sim \mathcal{N} \left( \frac{\sum_{t=2}^{T} \zeta(t) \zeta(t-1)}{\sum_{t=2}^{T} (\zeta(t-1))^2}, \frac{\sigma^2_{\zeta}}{\sum_{t=2}^{T} (\zeta(t-1))^2} \right). \tag{4.81} \]
### 4.D Tables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Post. Mean</th>
<th>95% Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>0.9</td>
<td>0.7572981</td>
<td>[0.3525789, 0.9008296]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1e-04</td>
<td>0.00241573</td>
<td>[0.0003270937, 0.01244498]</td>
</tr>
<tr>
<td>$\xi$</td>
<td>10</td>
<td>12.66894</td>
<td>[4.192819, 20.92908]</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.25</td>
<td>0.2109136</td>
<td>[0.1834899, 0.2765246]</td>
</tr>
<tr>
<td>$\alpha_x$</td>
<td>1.5</td>
<td>1.531373</td>
<td>[1.484182, 1.594037]</td>
</tr>
<tr>
<td>$\alpha_y$</td>
<td>1.2</td>
<td>1.16662</td>
<td>[1.111364, 1.222882]</td>
</tr>
<tr>
<td>$b_x$</td>
<td>-7</td>
<td>-7.274216</td>
<td>[-7.574485, -7.028051]</td>
</tr>
<tr>
<td>$b_y$</td>
<td>-8</td>
<td>-7.909817</td>
<td>[-8.213925, -7.631056]</td>
</tr>
</tbody>
</table>

Table 4.1: Model 1 (Near hard-core simulation, random number of events at each time period): Displaying the true values and Bayesian estimates of the model parameters. We run 20000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 50000 iterations of the Birth-Death algorithm.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Post. Mean</th>
<th>95% Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>1</td>
<td>1.162929</td>
<td>[0.07307848, 2.717557]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.5</td>
<td>0.9354644</td>
<td>[0.4215883, 1.764035]</td>
</tr>
<tr>
<td>$\xi$</td>
<td>500</td>
<td>496.1169</td>
<td>[486.1799, 507.3909]</td>
</tr>
<tr>
<td>$\Sigma_{1,1}$</td>
<td>0.1</td>
<td>0.3725116</td>
<td>[0.04201639, 1.685708]</td>
</tr>
<tr>
<td>$\Sigma_{1,2} = \Sigma_{2,1}$</td>
<td>0</td>
<td>0.06471913</td>
<td>[-0.5318637, 0.8403291]</td>
</tr>
<tr>
<td>$\Sigma_{2,2}$</td>
<td>0.1</td>
<td>0.4457842</td>
<td>[0.06186671, 1.857859]</td>
</tr>
</tbody>
</table>

Table 4.2: Model 2 (Simulation, random number of events at each time period): Displaying the true values and Bayesian estimates for some of the model parameters. We run 20000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 50000 iterations of the Birth-Death algorithm.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Post. Mean</th>
<th>95% Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>1</td>
<td>0.9988396</td>
<td>[0.9687213, 1.009379]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0</td>
<td>0.003991578</td>
<td>[0, 0.02398497]</td>
</tr>
<tr>
<td>$\Sigma_{1,1}$</td>
<td>0.1</td>
<td>0.2573227</td>
<td>[0.0955247, 0.4672203]</td>
</tr>
<tr>
<td>$\Sigma_{1,2} = \Sigma_{2,1}$</td>
<td>0</td>
<td>0.01105712</td>
<td>[-0.04812283, 0.1451561]</td>
</tr>
<tr>
<td>$\Sigma_{2,2}$</td>
<td>0.1</td>
<td>0.4169322</td>
<td>[0.09958883, 1.145102]</td>
</tr>
</tbody>
</table>

Table 4.3: Model 2 (Simulation of a hard-core model with fixed number of events at each time period): Displaying the true values and Bayesian estimates for some of the model parameters. We run 30000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 50000 iterations of the Birth-Death algorithm with change moves only. Note that $\xi$ is not affecting the number of points in this case ($n = 25$, fixed).
Table 4.4: Model 3 (Simulation with fixed number of events at each time period): Displaying the true values and Bayesian estimates for some of the model parameters. We run 10000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 15000 iterations of the Birth-Death algorithm with change moves only. Note that $\xi$ is not affecting the number of points in this case ($n = 25$, fixed).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Post. Mean</th>
<th>95% Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>4</td>
<td>4.094366</td>
<td>[4.040976, 4.112459]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.01</td>
<td>1.943212e-05</td>
<td>[2.570093e-06, 0.0003957944]</td>
</tr>
</tbody>
</table>

Table 4.5: Tornadoes Application, Model 1: Displaying the Bayesian estimates of the model parameters along with 95% credible sets. We run 15000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 50000 iterations of the Birth-Death algorithm.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Post. Mean</th>
<th>95% Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>23.8104</td>
<td>[20.37702, 27.70137]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.9664598</td>
<td>[0.9652222, 0.9674125]</td>
</tr>
<tr>
<td>$\xi$</td>
<td>611.9511</td>
<td>[605.574, 621.8022]</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>10.82354</td>
<td>[10.78444, 10.86611]</td>
</tr>
<tr>
<td>$\alpha_x$</td>
<td>-0.1218602</td>
<td>[0.04759888]</td>
</tr>
<tr>
<td>$\alpha_y$</td>
<td>-0.1110907</td>
<td>[0.2843646]</td>
</tr>
<tr>
<td>$b_x$</td>
<td>-92.60087</td>
<td>[-93.10514, -92.10223]</td>
</tr>
<tr>
<td>$b_y$</td>
<td>36.33277</td>
<td>[35.85291, 36.81489]</td>
</tr>
<tr>
<td>$p_1(F2)$</td>
<td>0.7048503</td>
<td>[0.6687929, 0.7405099]</td>
</tr>
<tr>
<td>$p_2(F3)$</td>
<td>0.2440887</td>
<td>[0.2114171, 0.2781614]</td>
</tr>
<tr>
<td>$p_3(F4)$</td>
<td>0.04629898</td>
<td>[0.03111901, 0.064273]</td>
</tr>
<tr>
<td>$p_4(F5)$</td>
<td>0.004761964</td>
<td>[0.001018164, 0.01158546]</td>
</tr>
<tr>
<td>Parameter</td>
<td>Post. Mean</td>
<td>95% Credible Set</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>------------------</td>
</tr>
<tr>
<td>$\rho$</td>
<td>2.206598</td>
<td>[0.2071125, 4.490314]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1.087929</td>
<td>[0.0444709, 2.855626]</td>
</tr>
<tr>
<td>$\xi$</td>
<td>600.8459</td>
<td>[589.3724, 620.2269]</td>
</tr>
<tr>
<td>$\Sigma_\phi(1,1)$</td>
<td>0.1963323</td>
<td>[0.03715305, 0.8345191]</td>
</tr>
<tr>
<td>$\Sigma_\phi(1,2) = \Sigma_\phi(2,1)$</td>
<td>-0.01305822</td>
<td>[-0.3962457, 0.2412033]</td>
</tr>
<tr>
<td>$\Sigma_\phi(2,2)$</td>
<td>0.1887883</td>
<td>[0.03502229, 0.8998766]</td>
</tr>
<tr>
<td>$\text{vec}(\Phi)[1]$</td>
<td>0.001882154</td>
<td>[-0.6522803, 0.6506615]</td>
</tr>
<tr>
<td>$\text{vec}(\Phi)[2]$</td>
<td>0.01229085</td>
<td>[-0.6831995, 0.8481349]</td>
</tr>
<tr>
<td>$\text{vec}(\Phi)[3]$</td>
<td>0.00667</td>
<td>[-0.6319787, 0.6357665]</td>
</tr>
<tr>
<td>$\text{vec}(\Phi)[4]$</td>
<td>0.05196297</td>
<td>[-0.608515, 0.8058517]</td>
</tr>
<tr>
<td>$\phi_{0,x}$</td>
<td>0.009739399</td>
<td>[-0.5802708, 0.6231767]</td>
</tr>
<tr>
<td>$\phi_{0,y}$</td>
<td>-0.0003664433</td>
<td>[-0.6077317, 0.6204879]</td>
</tr>
<tr>
<td>$\phi_{1,x}$</td>
<td>-0.06991322</td>
<td>[-1.201769, 0.727035]</td>
</tr>
<tr>
<td>$\phi_{1,y}$</td>
<td>0.08255069</td>
<td>[0.03502229, 0.8998766]</td>
</tr>
<tr>
<td>$\phi_{2,x}$</td>
<td>0.03821898</td>
<td>[-0.8578946, 1.062613]</td>
</tr>
<tr>
<td>$\phi_{2,y}$</td>
<td>0.04794142</td>
<td>[-0.6522803, 0.6506615]</td>
</tr>
<tr>
<td>$\phi_{3,x}$</td>
<td>0.02601216</td>
<td>[-0.948285, 1.057691]</td>
</tr>
<tr>
<td>$\phi_{3,y}$</td>
<td>0.173592</td>
<td>[-0.6481966, 1.754695]</td>
</tr>
<tr>
<td>$\phi_{4,x}$</td>
<td>0.0609904</td>
<td>[-0.8846516, 0.972938]</td>
</tr>
</tbody>
</table>

Table 4.6: Tornadoes Application, Model 2: Displaying the Bayesian estimates of the model parameters along with 95% credible sets. We run 10000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 10000 iterations of the Birth-Death algorithm.
| \( \phi_{4,y} \) | 0.1031962 | -0.8449322, 2.244293 |
| \( \phi_{5,x} \) | -0.002851645 | -1.121887, 0.8612827 |
| \( \phi_{5,y} \) | 0.1582667 | -0.8342418, 2.314868 |
| \( \phi_{6,x} \) | 0.03765363 | -0.7490769, 0.8272211 |
| \( \phi_{6,y} \) | 0.01013015 | -0.7048276, 0.7954374 |
| \( \kappa_1[0] \) | 0.3414294 | [0.02779556, 1.525342] |
| \( \kappa_1[1] \) | 0.04468772 | [0.01511415, 0.1036496] |
| \( \kappa_1[2] \) | 0.02381297 | [0.009341793, 0.04928505] |
| \( \kappa_1[3] \) | 0.01602525 | [0.006265732, 0.03359748] |
| \( \kappa_1[4] \) | 0.01254888 | [0.004560353, 0.02729344] |
| \( \kappa_1[5] \) | 0.01102656 | [0.003536779, 0.02568807] |
| \( \kappa_1[6] \) | 0.01082119 | [0.002909896, 0.02836776] |
| \( \kappa_2[0] \) | 0.3440752 | [0.02588816, 1.467024] |
| \( \kappa_2[1] \) | 0.0459065 | [0.01365316, 0.1053498] |
| \( \kappa_2[2] \) | 0.02417285 | [0.008743625, 0.05100296] |
| \( \kappa_2[3] \) | 0.01637742 | [0.006071461, 0.03448385] |
| \( \kappa_2[4] \) | 0.01276702 | [0.004400082, 0.02806576] |
| \( \kappa_2[5] \) | 0.01120546 | [0.003431507, 0.02639776] |
| \( \kappa_2[6] \) | 0.01106949 | [0.002817455, 0.02928837] |
| \( \zeta_0 \) | -0.01326444 | [-2.050802, 2.036076] |
| \( \zeta_1 \) | -0.02301747 | [-1.37479, 1.326716] |
| \( \zeta_2 \) | -0.03126289 | [-1.183709, 1.081311] |
| \( \zeta_3 \) | -0.0437938 | [-1.174834, 1.04973] |
| \( \zeta_4 \) | -0.05245246 | [-1.261148, 1.097873] |
| \( \zeta_5 \) | -0.05941277 | [-1.307846, 1.145974] |
| \( \zeta_6 \) | -0.06389189 | [-1.389118, 1.242211] |
| \( p_1(F^2) \) | 0.7050393 | [0.6691988, 0.7393965] |
| \( p_2(F^3) \) | 0.2439072 | [0.2115756, 0.2779321] |
| \( p_3(F^4) \) | 0.04626048 | [0.03131752, 0.06383538] |
| \( p_4(F^5) \) | 0.00479299 | [0.0009808832, 0.01155935] |

Table 4.7: Tornadoes Application, Model 3: Displaying the Bayesian estimates of the model parameters along with 95% credible sets. We run 10000 iterations of the MCMC (10% burn-in period), and at each iteration we sample the auxiliary point pattern using 10000 iterations of the Birth-Death algorithm.
4.E Figures

Figure 4.1: Model 1 Simulation (near hard-core model): displaying all events from time 1 to 4.
Figure 4.2: Model 1 Simulation (near hard-core model): displaying all events from time 5 to 10.
Figure 4.3: Model 2 simulation of a MSTPPP with random number of events at each time period. All the events are shown together in the top-left plot.
Figure 4.4: Model 2 simulation of a hard-core MSTPPP model with fixed number of events at each time period. All the events are shown together in the top-left plot. Note that within each time period, the events are further than $\rho = 1$ units away (hard-core distance).
Figure 4.5: Model 3 simulation with a fixed number of events at each time period. All the events are shown together in the top-left plot.
Figure 4.6: Tornadoes Application: displaying all tornado events from 2005 to 2010. The tornado power is indicated by a disc of different size.
Chapter 5

Future work and Conclusions

5.1 Extensions to the models for independent events

In Chapter 2 we presented a hierarchical Bayesian framework for modeling the intensity function of a DSTPP using finite mixtures for the location process and the one dimensional Hawkes Process for the temporal events via conditioning on time, i.e., $\lambda(s, t) = \lambda \lambda(t)$ $\lambda(s|t)$, without considering any covariate information. However, the inclusion of covariate information, such as the magnitude of the earthquakes could be important. In order to include covariates/marks one needs to model the DSTPP density based on $\lambda(s, t, \xi)$, where $\xi$ is the covariate (e.g., the earthquake magnitude). This modification will require new development for the forms of conditioning of the intensity function and will be investigated elsewhere.

In Chapter 3, we similarly presented a hierarchical Bayesian framework for modeling the intensity function of the DSTPP using finite mixtures for the location processes and the one dimensional Hawkes Process for the temporal events, via conditioning on location (i.e., $\lambda(s, t, \xi) = \lambda\lambda(s)\lambda(t|s)\lambda(\xi|s, t)$, where $\xi$ is some mark. As we have seen, this is only one of the possible conditioning forms. Since many of the parameters of the temporal process are conditioned on the location, their a priori models need to be defined over all possible locations in the window of observation, i.e., their priors are modeled via spatial random fields. The choice of the type of the random field is critical. In particular we use the one dimensional Hawkes process whose decay parameter is difficult to model, and we use a
simple exponential decay function depending on location that was modeled using a Gaussian process. However, choosing an alternative random field, such as the \( \chi^2 \)-random field may be appropriate. Similarly, certain model parameters can be modeled via random fields, in particular, the variance of the error process we use to evolve the mark probabilities; that is, instead of \( \sigma^2 \), use \( \sigma^2(s) \). In the models of the intensity function of the DSTPP discussed in Chapters 2 and 3 we did not consider the aspect of testing how well did the models perform in forecasting future states of the process, and this will be considered elsewhere.

Another future extension of this work will be using the birth-death procedure which circumvents the need to estimate the number of components for the mixtures as used in chapters 2 and 3. The purpose of dynamic modeling is to provide a means by which one can be able to do forecasting, however this was not considered in this context, mainly because of the high unpredictability of the Hawkes process. One possible solution to this problem, is the introduction of space-time varying covariate information, such as the existence of faults and other subterranean geo-dynamic factors that influence the occurrence of earthquakes. These might inform in the parameters model and aid in forecasting future states of the process.

5.2 Extensions to the models for interacting events

In Chapter 4 we presented a hierarchical Bayesian framework for modeling the joint distribution of a marked Markov space-time point process model, while taking care of potential interactions between the events. The three models presented can be used in different situations depending on the type of data under consideration. In these three models some parameters were independent of time (i.e., not dynamically evolving), in particular the parameters \((\rho, \gamma, \xi, p, \beta)\) and the marks were assumed to be independent of time. These parameters are responsible for important aspects of the models, namely, \(\rho\) is responsible for the distance of interaction between events, \(\xi\) determines the number of events, \(\gamma\) is responsible for the nature of interaction ("inhibition" or "attraction"), \(p\) is the parameter for the mark distribution, while \(\beta\) controls the effect of the covariate information at the locations of the
events. It is evident that most of the time these aspects of the models are dependent on
time, and thus, it is critical that these parameters be modeled using some evolution models,
and this extension will be considered elsewhere.

5.3 Concluding remarks

In this thesis we presented frameworks for modeling point patterns that occur in many
different fields of study by utilizing the inherent flexibility, clarity and functional utility of
the hierarchical modeling framework, together with the benefits of Bayesian methods as a
tool for analysis and inference.

In Chapter 2 we presented the hierarchical Bayesian framework for modeling the intensity
function of the DSTPP. The conditioning considered in this chapter is on the time stamps
(occurrence times of the spatial events). This methodology provides a framework that can
be thought of an extension of Poisson point process models that can include a time stamp
(discrete or continuous, in this exposition continuous). Using the Hawkes process, which is
itself a Poisson Point process, allows us to model and obtain the time stamps. More impor-
tantly, by using the Hawkes process we adhere to the natural ordering of time and we can
obtain an analytical solution of $\Lambda(T)$. The latter is typically approximated using numerical
methods and thus the results obtained in this chapter are exact and no approximations in
the model likelihood are required. As a result, we are able to achieve great reduction in
computational time in the iterations of the MCMC.

By extending the conditioning approach of Micheas et al. (2014) we are able to model
the space-time Poisson point process intensity $\lambda(s,t)$ in such a way that we allow replications
in time, by treating the times of events as a mark. Furthermore, we considered another
approach in the multivariate spatio-temporal model, which equally accommodates replica-
tions, with a major difference being that we treat the whole point pattern at a given time
as a mark. It is evident from the results in Chapter 2 from both the simulated data and
the application to earthquake data, that the models presented therein provide an appeal-
ing framework to model and capture the dynamics of a spatio-temporal point process. It is worth noting that the dimension of the parameter space changes across time because of the varying number of mixture components, which increases model complexity. Evolving both the truncated-Poisson parameter for the number of components and the mean for the Dirichlet parameter helps us in circumventing the varying dimension. In contrast to using infinite mixtures, the model presented provides summary statistics for different features of the intensity surface (mixture deconvolution), thus providing a means by which inference can be made about a specific feature that may be of interest. For example, the mean of a specific mixture component would be of interest in certain applications, such as, finding epicenters of major earthquakes, crime hot-spots in a city, allocation of emergency centers for ambulance demand, or centers for disease outbreaks, to name a few. Therefore, mixture deconvolution can help policy makers to efficiently allocate resources when faced with potential loss of life and property.

In Chapter 3, we similarly presented hierarchical Bayesian framework for modeling the intensity function of the DSTPP, via conditioning on location. The model considered there can be thought of as a variant of the multivariate space time model of Chapter 2, with the major difference being the form of conditioning.

Finally, in Chapter 4 we presented a hierarchical Bayesian framework for modeling the joint distribution of a marked Markov space-time point process model, while taking care of potential interactions between the events. The three models presented can be used in different situations depending on the type of data under consideration. The approaches presented in this chapter form the basis for modeling different kind of scenarios, for instance model 1 allows the main effect means to evolve over time, but none of the other parameters are evolving dynamically. This model would be appropriate in modeling fast evolving locations of the events, for example, a herd of animals as it moves fast across the domain of observation, with the animals staying close to each other (i.e., fairly clustered). Model 2 allows the main effects to evolve over time, while keeping the covariance matrix (the layout of the locations) fixed across time, which can be useful in modeling all the point patterns under model 1, but
in addition, can capture a more complicated variability structure. While, model 3 can be thought of as an extension of models 1 and 2, model 3 allows the covariances to evolve over time. These models can be particularly useful when modeling events with interactions and layout structure that varies over time. This can be achieved by varying the values of specific model parameters. For instance, in all three models, the value of the parameter \( \gamma \) can be used to model different scenarios. For example, a hard-core model for \( \gamma = 0 \), an inhibition model for \( 0 < \gamma < 1 \), a Poisson pattern for \( \gamma = 1 \) (no interaction effects) and finally a model for clustering for \( \gamma > 1 \). Unlike classic models such as the Strauss model, the proposed model has a valid joint distribution in the case \( \gamma > 1 \), which is an improvement in models with Strauss like-interaction functions used to model point patterns exhibiting clustering.
Bibliography


Vita

Justin Obwoge Okenye was born in Kisii, Kenya. Obwoge was admitted to Egerton University, Njoro, Kenya in 1990 and received a Bachelor of Education (Science) with a major in Mathematics in November 1994. From the May of 1995, he attended Egerton University, Njoro, Kenya and received a M.Sc. degree in Mathematical Statistics in November 1998. From Fall 2011 Obwoge joined the University of Missouri-Columbia as a Fulbright scholar and received a Ph.D. in Statistics in July 2021.

Obwoge married Gladys Moraa in August, 1998 and they are blessed with two issues Lynn Kemunto (Magokoro) and Ian Omambia (Tata moke). He plans go back to Kenya and serve as teacher of statistics at his alma-mater Egerton University, Njoro, Kenya.